

## 簡單與複雜型企業內網路下資料發掘的架構

# Simple and Complex Data Mining Frameworks in the Intranet Environment

楊亨利 教授

Heng-Li Yang

Professor

金士俊 碩士

Shih-Chun Chin

M.S.

國立政治大學資訊管理系

Department of Information Management, National Cheng Chi University

### 摘 要

近幾年來，企業內網路在各大企業中蓬勃發展，因此而產生在企業內網路上尋找資訊與知識的問題。對於許多歷史悠久、經營範圍廣泛的企業來說，要從其企業內網路上大量的文件中找出真正有用的、潛在的資訊與知識，誠非易事。因此有待以較深入的思考構面對於各企業的企業內網路架構進行檢視，才能找出適當的資料發掘作法與技術，以解決此一難題。本研究先對企業內網路、資料發掘等文獻與現況進行整理分析，並提出與企業內網路複雜程度相關之參考思考構面，進一步，定義簡單型與複雜型的企業內網路。然後，以 Han、Zaine 與 Fu (1995) 之概念樹及多層次資料庫的觀念，採用我們以往研究成果中對於網際網路資料發掘所提出的理論架構與作法（楊亨利、金士俊，民 90 年），特別著重企業在授權與知識視野上的思考，提出簡單型與複雜型企業內網路下對資料內容發掘的架構，設計適用的發掘系統。並部份實作其雛形，以驗證其可行性。

**關鍵字：**資料發掘、知識發現、企業內網路、知識管理

### **Abstract**

In recent years, with the widespread use of the Intranet, Intranet data searching has become an important but problematic issue. For the enterprises that have been established for a long time or have widely operated on different locations, it is an extremely demanding task for them to identify valid and potentially useful patterns from the huge amount of documents in the Intranet. In order to solve this problem, it is essential to take a close and careful examination of different Intranet frameworks, so that appropriate data mining approaches and techniques can be worked out. This research began with a literature review of the fields of Intranet applications and data mining. Based on the review of current theories and practices, the second part of the research focused on necessary dimensions for evaluating the complexity of Intranet systems. In the third part of the research, the frameworks for simple and complex Intranet data mining systems were then proposed, three types of data mining approaches were designed and a prototype was developed. The ideas of “concept hierarchy” and “multiple layered database” (Han, Zaine and Fu, 1995) and Yang & Chin’s (2000) approach of Internet data mining were adopted. Besides, this research also paid particular attentions to the issues of “authorization” and “knowledge view” when designing the frameworks and developing the prototype. In the last part of the research, the feasibility, practicality and potential uses of our approaches and prototype were discussed, and some directions for future research were suggested.

**Keywords :** Data Mining, Knowledge Discovery, Intranet, Knowledge Management

## 壹、導論

對一個企業體而言，若能有效地自所有資料來源中找出比競爭者更多的未知知識與資訊，自然就更有致勝的機會。自從全球資訊網(World Wide Web，簡稱 WWW 或 Web)發展以來，由於其高度整合超鏈結與多媒體展示的超媒體能力，帶動進一步網際網路(Internet)的應用高潮。許多企業也開始使用網際網路科技來主導企業內的商業事務，造就一波建置企業內網路的熱潮 (Intranet Group of NASA, 1995)。這使得企業內網路中以超文字格式作為主體的資料成為企業重要資料來源，如何由其中找出未知的知識與資訊，如何將資料發掘技術應用在企業內網路的架構，就有進一步探討的必要。

若與傳統資料庫相較，企業內網路下的資料發掘面臨以下四項課題必須解決：

(1) 如何找出適用的思考構面，用以客觀分析與評估不同企業間企業內網路架構的差異。(2) 如何就各種不同的企業內網路架構特徵，規範出能夠在不同企業內網路架構環境下對超文字格式資料進行資料發掘的模式。(3) 如何能在不同企業內網路架構下將作為企業內網路資料主體的超文字格式資料上附加適合的結構化包裝，增加對其進行資料發掘的價值與意義。(4) 如何能夠在不同的企業內網路架構下利用適當的演算法與查詢界面，輔助使用者逐步正確表達其真實意圖，並且利用改寫規則的相關知識，

過濾無意義的查詢，縮小其真實意圖的範圍，作有效的資料發掘。

上述的最後一個課題，係屬智慧型資料庫界面研究領域，不在本研究範圍之內。除此之外，針對前三個的課題，本研究希望藉由文獻整理，探討企業內網路架構下的資料發掘，並建構進行資料發掘的相關架構，最後發展雛型系統，以達成以下三項目的：

(1) 利用對於資訊科技思考構面與構型的探討，提出對於思考企業內網路相關問題有所助益的構面，用以解讀不同組織之企業內網路環境以及作為在該環境發展相關應用之參考。(2) 建立不同企業內網路可行的資料發掘架構，讓其中的非結構化超文字格式資料在合適的架構下，經有意義的結構化格式資料包裝儲存後，可擷取出知識，發掘所需之相關資料。(3) 發展企業內網路資料發掘雛形，以驗證相關論述的有效性。

## 貳、文獻探討

本節將探討與本研究相關的一些文獻。

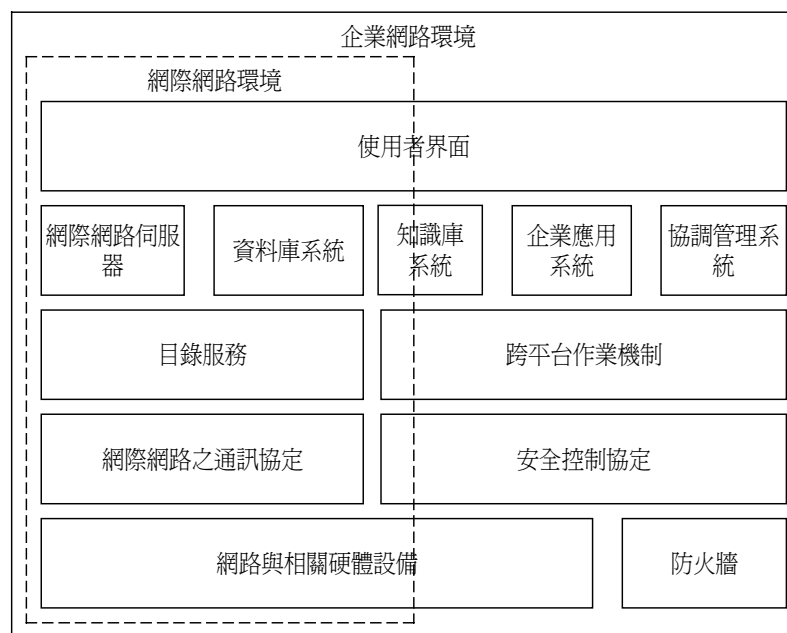
### 一、企業內網路

雖然學者間對於企業內網路的定義一直存在許多不同的看法，但大部分看法中基本上同意企業內網路是一種企業應用網際網路的技術與標準所架設的網路環境，晚近的學者 (Kalakota and Whinston 1997；Tellen 1997；陳建安 民 87) 則強調全球資訊網的技術與標準是為企業所引用的網

際網路標準中，比較重要的一環。在本研究中，定義企業內網路就是企業利用網際網路的技術，特別是全球資訊網的技術與工具來達成企業或組織內部的資訊需求所建置的資訊系統環境。它以網際網路的電腦標準協定、軟體、網路設備作為基礎；並針對企業組織內部對於資訊安全、控制、流程等相關需求作電腦標準協定、軟體、網路設備等相關的改革與延伸。

在對企業內網路的了解上，本研究參考了國內學者陳年興與陳建安提出之企業內網路的功能架構（陳建安、陳年興，民 86）。在其架構下，企業內網路乃是以網際網路為基礎所發展出來以符合企業內部使用的網路環境。企業內網路與網際網路兩者最大的差異點乃是因為前者主要供給企業內部使用，必須較後者在安全性上有較好的控制與管理性。陳建安與陳年興的架構常為相關研究者所引用。它具備以下四個優點：(a)可以釐清網際網路與企業內網路環境的

差異、(b)強調整合與跨平台的特性、(c)訂定相關功能區塊的階層關係、(d)可以解決企業與傳統主從架構的相關問題。但是其架構仍有以下三個缺失：(a)並非全部的網路及其相關硬體設備均為網際網路與企業內網路環境所共用(企業內網路亦具備專屬之網路及其相關硬體設備)。(b)未能表現防火牆作為企業內網路屏障的功能角色。(c)企業內網路中的使用者界面設計可能因為作業需求的不同，而和網際網路有所區別。因此本研究提出一個改良架構，如圖 1 所示。該圖表示了兩個主要部份：一是現有的網際網路環境（虛線範圍內），另一個則是企業網路獨有的部份（虛線範圍以外，實線範圍以內），此二者共同組成整個企業網路環境。基本上，圖 1 仍援用陳建安與陳年興（民 86）圖例，但虛線包圍範圍不同，並將防火牆特別從網路及其相關硬體設備中獨立出來，以凸顯其功能角色。



資料來源：本研究

圖 1 企業內網路功能架構修正圖

## 二、不同組織的資訊科技運用

對於不同組織下資訊科技運用的分類與整理，是在 1980 晚期才開始被提出，而加以討論。有一些學者採用事先分類的方法，將資訊系統歸類入一些預設的群組中 (Leifer, 1988; Ahittuv, Neumann and Zvirin, 1989)。而另外一些學者則將資訊系統利用一些維度加以分群分析，演繹出不同的資訊科技結構分類 (Fielder, Grover and Teng, 1996)。在這些討論中，對於組織結構除了使用傳統的功能型組織、產品型組織以及矩陣型組織的分類方式外，Mintzberg (1979) 的組織構形也常被廣泛採用 (Leifer, 1988; 楊亨利、林俊宏, 民 86)。

Leifer (1988) 採用 Mintzberg 的組織構形理論，但將事業部依照事業部彼此間的合作關係區分為 A、B 兩型，而區分出六類組織結構構形，並將現行資訊系統區分為以下四類：(1) 獨立式系統 (Stand-alone Systems)，由分散隔離的電腦所組成；(2) 集中式系統 (Centralized Systems)，由隔離的中央處理器與終端機所組成；(3) 分散式系統 (Decentralized Systems)，由網路連結相同等級的電腦所組成；(4) 分權式系統 (Hub-and-spoke Systems)，由透過網路連結中央處理器與其他電腦所組成。

Fielder、Grove 與 Teng (1996) 則認為資訊系統的分類方式不應該預設分類結果，而應先找出分析的維度，提出合理的假設，再利用分群分析演繹出適當的分類，他們使用三項分析維度：(1) 運算集中的程度 (The Extent of Computing Process)、(2) 電腦支援通訊的等級 (The Degree of Computers Support Communication)、(3) 電腦分享資料與應用程式的能力 (The Ability of Computers to Share Data and Application Programs)。由此三個維度，經過分群分析與相關統計檢定後，演繹出以下四個的資訊科技結構分類，並產生對於組織結構形態之間的對應：(1) 集權式運算 (Centralized Computing)，集中處理、低通訊、低分享的資訊科技結構，對應於低整合與決策形成集中的功能型組織的企業；(2) 分權式運算 (Decentralized Computing)，不集中處理、低通訊、低分享的資訊科技結構，對應於決策形成不集中的產品型組織的企業；(3) 集權式合作運算 (Centralized Cooperative Computing)，集中處理、高通訊、高分享的資訊科技結構，對應於高整合的功型組織的企業；(4) 分權式合作運算 (Decentralized Cooperative Computing)，不集中處理、高通訊、高分享的資訊科技結構，對應於高整合的矩陣型與產品型組織的企業。

國內學者楊亨利與林俊宏等人(民 86) 則根據 Mintzberg 所提出的組織構形以及 Jarvenpaa 與 Ives 在 1992 年所提出的科技構形的概念，以更廣泛的層面上探討資訊構形

與組織構形之間的關係，整理出七種資訊科技構面如下：資訊部門主管地位與在組織層級、硬體的分散、資料庫的散佈、系統操作與維護、系統開發、資訊系統規劃、以及系統應用層次。另一方面，他們進而修改 Leifer 的四種資訊系統，提出簡單型、控制型、支援型、供應商整合型、與溝通樞紐型成爲五種組織內的資訊科技構形，並使用前述的七種資訊構面區分其不同。

### 三、企業內的知識

在充滿不確定的經濟體系中，知識管理是一個重要而值得討論的議題。知識是一種流動性質的綜合體，其中包括結構化的經驗、價值、以及經過文字化的資訊。此外，知識也包含專家獨特的見解，其可爲新經驗的評估、整合與資訊等提供架構。知識起源於智者的思想，在組織中，知識不僅存在文件與儲存系統中，也蘊涵在日常例行工作、過程、執行與規範當中。知識和資訊不同，資訊是一種訊息，它通常是透過文字傳送，但知識比資訊更接近行動。因此知識的評估應該以它對決策或是行動造成的影響爲準 (Davenport and Prusak, 1998)。

當然，對於任何企業而言，知識並非可以平白產生的，而必須以一定的代價，透過不同的方式來取得，常見的方式包括下列幾種：內部發展、借助外部協助內部發展、市場採購、廠商間合作、與購併 (Helleloid and Bernard, 1994)。不論以何種方式所取得的知識，不經過整理的步驟，企業便無法將

之納入企業之中進行管理。進行知識整理的四大原則如下：(1) 決定知識整理的目的何在；(2) 在眾多不同形式的知識中，掌握適合的知識來達成目標；(3) 評估知識，判斷是否有用或者合適整理；(4) 整理者必須找出適當的媒介來進行整理和傳播 (Davenport and Prusak, 1998)。

取得的知識在經過整理並納入管理之後，必須透過企業的成員進行轉換與學習的過程，才能將之轉化為該企業的核心能力。知識轉移包括兩大行動：傳達知識給潛在的接受者、其由該接受的個人或團體加以吸收；如果知識未經接受者吸收，就不算是真的轉移成功。知識轉移的目的是改善組織行事的能力，並進而提昇其價值。所以，知識轉移 = 傳達 + 吸收 + 利用 (Davenport and Prusak, 1998)。知識常被區分為內隱和外顯兩種 (Nonaka and Takeuchi, 1995)。以此觀點，知識轉換可區分為四個模式：(1) 內隱至內隱 (共同化)、(2) 內隱至外顯 (外化)、(3) 外顯至外顯 (結合)、(4) 外顯至內隱 (內化) (Nonaka and Takeuchi, 1995)。

組織應進行知識蓄積以形成組織記憶，來累積其長期的知識基礎。而知識的蓄積需配合知識的類型進行。如以企業中顯性知識為例，蓄積顯性知識要用到四種基本的資源：(1) 顯性知識的資料庫或知識庫，(2) 累積、改進、管理、傳播知識的流程，(3) 組織內執行並管理改進流程的人員職位，(4) 支援上述資料庫或知識庫和各項流程的資訊技術 (Zack, 1999)。

組織在吸收知識時可透過組織的溝通系統(包括組織與外界的溝通、單位之內的溝通)及透過網路的效果 (Cohen & Levinthal, 1990)；即組織成員除個別擁有技術知識外，還必須知道組織內外何處有互補性知識的專家，以強化個人與組織的吸收能力。不同背景和知識的個人會影響組織的吸收能力，而組織的吸收能力主要根植於個別能力之間的連結，這種連結也就是所謂的知識分享。知識分享是一種溝通

(Hendriks, 1999)。當組織成員向他人學習知識時，就是在分享他人的知識，而「知識接收者」必須有重建的行為。因此，知識分享牽涉兩個主體：知識擁有者和知識需求者。而知識分享程序必須包含知識擁有者將知識「外化」(Externalization)與知識需求者「內化」(Internalization) 兩個步驟。

組織中的知識分享與組織中所存在的社會網路 (Social Network) 型態有密切的關聯，一般而言，社會網路可分為三種型態 (Brass and Burkhardt, 1992)：

(1) 工作流程網路：組織的任務是由許多工作者來共同完成，因此，工作流程中上下游的工作者的績效有部分是相互依賴的。工作者彼此交換投入與產出，若所投入的資源或可分配產出的選擇方案較多，就可以減少對他人的依賴程度，其權力也將隨之增加。這種建立在各部門分工基礎上、彼此相互依賴的工作者間，就形成了「工作流程網路」。

(2) 溝通網路：又可稱為「諮詢網路」。組織內的成員為解決工作上的問題或促進任

務的達成，會不時地做意見上的溝通、資訊的交換、知識與經驗的流通。因此，本研究又將此稱為「知識分享網路」。在溝通網路或知識分享的網路中，人們的互賴是建立在資訊交換基礎上。一個位處於溝通網路中央位置的成員，可能具有取得資訊（密切度）與控制資訊（中介度）之潛力，因而具有潛在的權力。

（3）友誼網路：組織中的成員會因彼此有相同的喜好或友誼的建立而產生連結。在這種網路中，資源或友誼不一定是權力的直接來源，但是借助友誼此一工具的幫助卻可能獲得資源（如：資訊、獎懲等），這使得友誼網路的中樞性與權力間具有正向關係，也就是說居於友誼網路較為中樞的人，在企業中獲取權力的機會就較高。情感連結是社會資本的一項來源（Friedman and Krackhardt, 1997）。人們傾向與自己相類似的人有互動的關係，由於彼此類似、共同之處較多，比較容易相互了解、溝通、接受、預測以及建立信任感，因而有較多的空間可以發展情感。

事實上，所有組織多少都存在以上三種社會網路，而且當某種社會網路的連結經過相當的強化，就有可能在企業內形成所謂的社群（Community），伴隨企業內網路的發展，依附於電子網路上的溝通網路與友誼網路，也很可能在企業內形成虛擬社群（Virtual Community）的特殊現象而擴大其在知識分享上的影響力（江姿慧，民 89）。

#### 四、資料發掘

資料發掘（Data Mining）或稱資料庫中的知識發現（Knowledge Discovery in Databases）是指由資料庫中選擇合適資料、資料處理、資料轉換、資料發掘至結果評估，以獲得「非顯然的」、「隱含的」、「前所未知的」、而「可能有用」資訊的過程（Frayyad, 1996），各種針對現行資料庫的資料發掘方法與技巧已經大量被提出，而廣泛地為人所討論（Frawley et al., 1991; Yoon and Kerschberg, 1993; Grupe and Owrang, 1995; Fayyad, 1996; 周立平，民 84 年；薛如芳，民 84 年）。

Han 的概念樹學習法（Han, Cai and Cercone, 1992）是近年來在資料發掘的研究中頗受重視的方法之一。其主要的精神是由領域專家依其對於領域的了解設立概念樹，於資料發掘過程中，對已建有概念樹者的資料屬性，依其概念樹不斷向上抽象化，直到抽象化之屬性個數符合預先設定之上限為止，再整理成邏輯規則。讓使用者所看到的，不是原始資料庫中零碎的資料，而是結合此領域的概念與術語而成的整體知識。目前關於概念樹學習法的探討，已經十分豐富，不但已經能解決由關連式資料庫中（Relational Database）發掘出特性規則（列出如研究生某特定族群的特性）（Dhar and Tuzhilin, 1993; Han, Cai and Cercone, 1993）、區別規則（比較如研究生與大學生多類族群的差異）（Agrawal, et al., 1992; Agrawal, Imielinski and Swan, 1993）、關聯規



則(指出事物間的關聯性,如買牛奶與買麵包的關聯性)(Agrawal and Skikant, 1995; Han and Fu, 1995; Houtsma and Swami, 1995)、分群規則(將資料庫依其特性分群)(Chen, Han and Yu,1996)、進化規則(表示資料在持續時間記錄下之變化趨勢)(Han, Cai, Cercone, and Huang,1995)的問題。同時也已經對於將概念樹應用於地理空間(Han, Koperski and Adhikary, 1996)、物件資料庫(Han, Nishio and Kawano, 1994)與非常大資料庫(Han and Fu,1995)有過許多討論。

## 五、網際網路上的資料發掘

隨著近年來全球資訊網的蓬勃發展,全球資訊網上的資料發掘(Data Mining on WWW)或全球資訊網之資料發掘(Web Mining)(Etzioni,1996; Zaine,1999)已成為一個挑戰性的研究課題。也有一部份的學者則將此一議題與過去舊有的網際網路資源(例如:Gopher、WAIS、BBS)上的資料與知識的尋找的研究相結合,稱之為網際網路的資源發現(Internet Resource Discovery)(Bowman, et al.,1994;Schwartz, et al. 1992)。

依照對於全球資訊網不同的資訊需求,全球資訊網之資料發掘(Web Mining)可以分成以下幾種不同的分類(Zaine, 1999):(1)對於全球資訊網資料內容的資料發掘(Web Content Mining),包括對於以HTML 語言格式資料為基礎的原始網頁資

料內容的資料發掘(Web Page Content Mining),或是以使用者利用搜尋引擎對於原始網頁作相關查詢所得結果的資料發掘(Search Result Mining)。(2)針對構成全球資訊網資料基礎的以HTML 語言超鏈結特性,分析網頁資料互相指向與參考的鏈結架構(Web Structure Mining)的資訊發掘。包含對於網頁指向參考別的網頁(或網址)、或是網頁被其他網頁(或網址)指向參考的兩種不同方向分析。(3)對於全球資訊網使用現狀的資訊發掘(Web Usage Mining)(Pei, et al., 2000; Zaine, Xin, and Han,1998; Chen, Park and Yu,1998; 陳仕昇等,民八十八年),包括對於被存取網站伺服器所記錄的全部使用者的使用記錄檔(Log)內的相關資訊(General Access Pattern Tracking)或是針對被存取網站伺服器所記錄的特定的使用者的使用記錄檔內的相關資訊(Customized Usage Tracking)作資料發掘。前者的目的在於對網站整體的架構作相關的調整,後者則有助於讓網站可針對不同使用者達到個人最佳化。

對全球資訊網的資料內容的資訊發掘若只使用搜尋引擎與目錄服務顯然不能完全解決超文字格式資料非結構化的問題,有時反而有時還會衍生出例如網路頻寬過度負荷等新的障礙。為了輔助搜尋引擎與目錄服務發揮應有功能與消除其衍生障礙,Han等人提出一個構想,他們希望個別超文字文件之提供者能遵循某些資料結構提供資料,而後進行抽象化之萃取,並將其結果儲

存為一些結構化的關聯式資料表格，再對這些表格進行搜尋與再抽象化之萃取；逐步構形成一多層狀之資料庫（Multiple Layered Database, MLDB）。當使用者希望由全球資訊網上的資料中找出資訊時，可以分別就其需要，直接對上層資料庫做相關的查尋（Han, Zaine and Fu, 1995），這樣就可以既快速又省事的得到所需的資料。如果一個這種全球資訊網的多層資料庫能成功建立，那麼就可以提供一個層級組織井然的全域性的觀點，讓使用者不必直接從漫無組織的全球資訊網去擷取資訊，Zaine 將這樣的觀點稱為一個虛擬的網際觀點（A Virtual Web View, VWV）（Zaine, 1999）。他並指出，近年來可延伸式標示語言（Extended Markup Language, XML）標準的逐漸成熟，會有助於這種虛擬觀點的成形。

Zaine 與 Han 進而提出一個全球資訊網資料發掘語言（WebML），企圖讓使用者能夠使用類似 SQL（Structured Query Language）的查詢語言語法，由前述的多層次資料庫中發掘想獲得的知識（Zaine, 1999）。他們已有部份的系統雛型實作，不過不論是在目前傳統超文字標示語言，或是在未來可延伸式標示語言的環境下，其 MLDB 中需要個別超文字文件之提供者所遵循的資料結構如何形成？除了原始資料外的上層資料結構應如何才能為大家所共同接受？在 Han 與 Zaine 的研究與實作中並沒有足以讓人滿意的答案。

針對以上困境，有人主張在超文字格

式的資料架構中加強其結構化之段落定義，使之具有網際網路資源描述與著錄意義內涵。企圖透過新的定義，使這些或由資料原作者填寫，或利用對非結構化本文分析之技術所產生的結構化段落，以解決超文字格式資料之非結構化的問題。其中，較出名的相關研究包括由 OCLC/NCSA Metadata Workshop 整理出十二項具有網際網路資源意義內涵的資料項目，遵循超文字文件目前之架構，擴充超文字語言較具結構性的 HEAD 標記段落中的 META 標記子段落（Raggett, 1997; Weibel, 1996; Weibel, Godby and Miller, 1996），稱為“Dublin Metadata Core Elements Set”。Desai（1997）則認為 Dublin Core 過於簡略與遷就現行架構，提出直接修正超文字語言之架構，徹底改寫其中較具結構性的 HEAD 標記段落，重新定義其子段落組成，而成為具有網際網路資源描述與著錄意義內涵的語意標頭（Semantic Header）。

在我們過去的研究（楊亨利、金士俊，民 90）中，延伸 Han 與 Zaine 的 VWV 架構並加入前述 Meta Data 的概念予以修正，曾提出一個全球資訊網上資料發掘之整體架構設計及運作方式。與 Han 與 Zaine 的 VWV 架構相較，我們過去所提出的架構主要的不同在於：

（1）提出一個涵蓋於全球資訊網實際文件一般屬性需求，而且具有足夠彈性，可以滿足全球資訊網上不同領域應用的 Meta Data 定義，作為一個可能為一般超文字文件提供

者所能接受與遵循的上層資料結構。

(2) 提出虛擬全球 Meta Data 資料庫的概念，利用現行存在的搜尋引擎與目錄服務的機制，具體成形一個可行的 MLDB 的組織架構。

(3) 提出在網際網路資料發掘的過程中，可以加入領域與改寫規則與專門字彙涵意的資料庫等智慧型系統的輔助，讓資料發掘的過程更有效率以及彈性。

## 參、企業內網路下資料發掘之概念性系統架構

### 一、企業內網路之思考構面

本研究先整合了前述文獻，形成六個企業內網路思考構面，如下所述：

(1) 企業內電腦系統網路連結的程度：企業是否具備相關的軟硬體設施，讓企業中的電腦可以透過內部的網路設備互相連結，傳送資料與訊息。

(2) 企業內網路上電腦系統運算集中的程度：在企業資訊系統中，負責相關運算的工作的電腦，與單純接收與顯示其他電腦傳送之運算結果的電腦的比例。

(3) 企業內網路上電腦間分享資料與應用程式的能力：在企業資訊系統中，資料與應用程式資源在電腦之間共享的程度。

(4) 企業內網路系統規劃與概念設計：包含企業內網路系統目標與公司策略之結合、企業內網路架構之發展、資料格式與通訊標準的訂定等策略性或技術面的相關決策。

(5) 企業內網路系統實際的建置與維護：企業內網路系統建置與維護之權限問題，攸關網路資訊系統品質之優劣。

(6) 企業內網路應用層次與相關的資料操作授權：企業內網路環境的使用是屬於策略、管理、或作業層次的企業活動？在企業內應用目的是屬於個人、工作群組、組織、或組織間？為達成此種應用層次，在資訊系統中給予相關人員的操作授權為何？

前三個構面係整合了 Liefer (1988) 之「網路連結」、「運算集中程度」的維度以及 Fielder、Grove 與 Teng (1996)「運算集中程度」、「電腦支援通訊等級」、「電腦分享資料與應用程式能力」的維度，還有楊亨利與林俊宏 (民 86 年) 七種資訊科技構面中關於「硬體分散」、「資料散佈」的構面等想法。第四構面來自楊亨利與林俊宏之「資訊系統規劃」構面，但著重於架構設計、協定標準的訂立與規範。第五構面乃依據楊亨利與林俊宏之「系統開發」構面，及「系統操作與維護」構面中關於系統維護的觀念。第六構面來自楊亨利與林俊宏之「系統應用層次」構面。至於楊亨利與林俊宏之「資訊主管在組織層級」構面，則因為本研究討論焦點侷限在企業內網路，而非著眼於整個資訊科技架構上，我們僅將其中部分觀念如企業對於資料操作授權等個別項目併入企業內網路應用層次構面。

此六構面中，前三個著重於描述資訊科技目前在企業中的實際架構，若配合圖 1 本研究所修正之企業內網路功能架構，可以

用於判斷企業內網路目前存在於企業之複雜程度。後三個思考構面著重於描述組織是如何應用資訊科技，可判斷企業實際應用企業內網路需要之複雜程度。若我們將一個「簡單的企業內網路」架構定義為企業單純的採用網際網路的技術作為企業資訊系統網路連結的標準。而一個「複雜的企業內網路」架構定義成企業為滿足特殊需求，例如大量不同的文件與應用程式的整合，延展網際網路的技術，大幅修改與增強網際網路欠缺的機制，使企業資訊系統呈現複雜的網路連結<sup>1</sup>。我們可對不同企業之前三、後三構面分別評分後，以加總後分數作為X座標與Y座標，可得到如圖2企業內網路的評估與分析的思考模型。

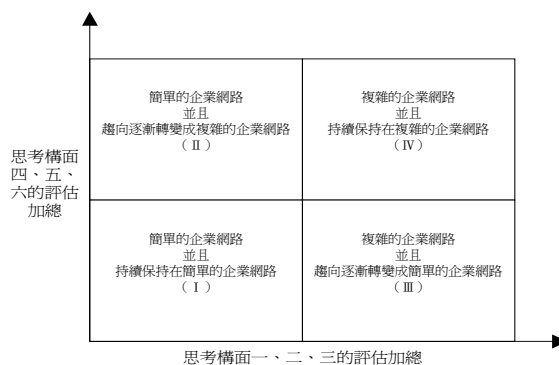


圖2 企業內網路的簡單與複雜評估模式

也就是說，企業內網路的複雜或簡單，若以六個思考構面作相關判斷時，將會呈現以下四種形態的一種：

(1) 形態 I：該種企業內網路目前在網路連結、資源分享與合作運算上的資訊科技僅採用現行的網際網路標準，而其企業之系統開發、應用層次與應用範圍的發展需求也相當單純。此種企業內網路是屬於簡單的架構，並且將持續呈現簡單的架構。

(2) 形態 II：該種企業內網路目前雖在網路連結、資源分享與合作運算上的資訊科技僅採用現行的網際網路標準，但其企業之系統開發、應用層次與應用範圍的發展需求並不單純，目前之網際網路資訊科技可能在未來（甚至目前已）完全無法滿足其實際需求。該種企業內網路目前雖呈現簡單架構，但是必定會逐漸朝轉變為複雜架構。

(3) 形態 III：該種企業內網路目前網路連結、資源分享與合作運算上的資訊科技已經採用延展網際網路的技術，修改與增強網際網路欠缺的機制，但其系統開發、應用層次與應用範圍的發展需求卻相當單純。這或許是網際網路標準在功能上的缺失逼使該企

<sup>1</sup>一個簡單型架構是指企業主要採用網際網路技術作為其資訊系統網路連結的標準，只對此標準進行小幅度安全與功能上的強化，如以圖1的功能架構來觀察，其功能區塊中網際網路所佔有的比重應相當高。而一個複雜型架構則是指企業為滿足特殊需求，例如，企業覺得以HTTP為基礎之瀏覽器所包含的HTML中Tag不符需求，另設計以SGML為基礎的瀏覽器；甚至，另在HTTP之通訊協定上加一層編碼，以應付廣大地區傳送中的可能安全或病毒問題；或覺得SSL之128bits標準太低，另設計256bits為基礎之安全機制。如以圖1的功能架構來觀察，其功能區塊網際網路所佔有的比重則相對上會比較低。

業作技術上的延伸，也有可能是企業內網路規劃不良、或遷就舊有系統而採行技術上的延伸。該種企業內網路目前雖是屬複雜的架構，但將會隨著網際網路標準的進步，或是企業使用內部網路經驗上的成熟，而逐漸調整至簡單的架構。

(4) 形態IV：該種企業內網路目前在網路連結、資源分享與合作運算上的資訊科技已經採用延展網際網路的技術，修改與增強網際網路欠缺的機制，而其企業之系統開發、應用層次與應用範圍的發展需求也相當複雜。該種企業內網路目前是屬於複雜的架構，並且將持續呈現複雜的架構。

綜合而言，在形態I與形態IV中，企業內網路的架構都已經達到其穩定的適當複雜程度，而且會繼續保持下去；存在有對應的簡單型與複雜型的企業內網路資料發掘架構可以遵循，也是以下本研究的探討的重點。而形態II與形態III仍屬變動中，所以其資料發掘的需求與限制都是不確定的，很難找出對應的簡單型與複雜型的企業內網路資料發掘架構可以遵循。不過，他們也可依本研究所提出之六個構面思考，就其在企業內網路所面臨的需求與限制，修改與形態I與形態IV對應的資料發掘架構，以求更適合其之作法。這正是本研究提出這些思考構面、模型與參考架構之目的所在。

## 二、與資料發掘相關的企業內網路之思考構面

在討論企業內網路下資料發掘的相關

問題時，我們主要根據的是前述六個思考構面的第二個、第三個以及第四個、第六個思考構面，其原因如下：

(1) 第一個思考構面主要作用在判斷一個企業，是否具備發展企業內網路環境之資訊科技基礎。在討論企業內網路下資料發掘的相關問題時，已經存在企業內網路可算是前提假設。即使是一個企業的連結程度不到百分之百，但就討論企業內網路時，顯然應就其以連結的部份來討論其企業內網路架構。所以此一思考構面的意義也就相形較小，而較不必特別予以考慮。

(2) 第五個思考構面主要作用在思考一個企業是如何去建置與維護其企業內網路環境的相關議題，但本研究主題是在何種的企業內網路環境下，有哪些有意義的資料可以發掘與該如何去從事資料發掘。所以我們特別著重在架構、文件格式、通訊以及其他協定等屬於規劃與概念設計層次的討論。若依這些討論所規劃與設計的企業內網路，即便採用不同的技術進行建置與維護，影響將只於效率與成本等較實務層面的議題，而不會影響到資料發掘工作本身的意涵。所以同樣地，此一思考構面的意義也相形較小，較不必特別予以考慮。

所以，以下是我們四個相關構面的思考：

(1) 電腦系統運算集中的程度：在簡單型的環境下，運算集中於少數伺服器上執行。所以當使用者有資料發掘的運算需求時，其運算將由一部伺服器完成，使用者的瀏覽器

只單純顯示此一運算的結果。而在複雜型的環境下，運算是分散在許多伺服器與使用者客戶端之間進行，必須使用一些比較複雜的延伸技術。當使用者有資料發掘的運算需求時，其運算工作將由使用者電腦分配給企業內不同的電腦執行，再由原使用者電腦彙整結果並顯示於其瀏覽器上。

(2) 電腦間分享資料與應用程式的能力：在簡單型環境下，企業資料通常集中保存在企業內少數的伺服器上。所以對於資料發掘而言，只需蒐集這些伺服器中的原始資料，並直接使用儲存在從事資料發掘運算的伺服器中其他資訊輔助，便可以完成相關工作。而在複雜型的環境下，無論是作為資料發掘對象的原始資料，或是用以輔助發掘工作之其他資訊，均可能散佈於組織內不同的電腦之中。因此必須使用特殊的存取機制，讀取與整合在不同的電腦中的資料或資訊，才能夠完成資料發掘的工作。

(3) 網路系統規劃與概念設計：在簡單型環境下，所有策略性或技術面的相關決策都相當單純。一切皆以網際網路通用的資訊科技與標準協定作為考量的範圍，例如以網際網路通用的 TCP/IP 協定作為通訊的基礎協定，以網際網路通用的超文字文件作為文件儲存的唯一格式。資料發掘的對象也就侷限在超文字文件，概念設計自然十分單純。而在複雜型環境下，策略性或技術面的相關決策就較為縝密而且融合企業內較多意見，所以必須採行一些延伸網際網路的資訊技術，甚至特別制定特別的標準或協定，例如

以通訊閘讓企業網路可兼容非 TCP/IP 的通訊協定、容許在企業所有瀏覽器使用特別的加掛 (Plug-in) 程式閱讀特殊文件格式等。資料發掘的對象可能是各種文件甚至非文件，所以概念設計必須較為複雜。

(4) 企業網路應用層次與相關的資料操作授權：簡單型下資料發掘在應用上偏向於只包含作業管制層面與部份較簡單的管理管制的運用，且若自組織階層來觀察時，此資料發掘系統由於發掘用的資料與輔助其他資訊較為有限且集中保存。所以雖然也可能被應用在個人、工作群組或組織的不同層次，但是為不同層次區分的資料操作授權機制卻不需太複雜。而複雜型下資料發掘在應用上包含各種作業管制、管理管制與策略制定層面的運用，且發掘用之資料與輔助資訊來源廣闊而分散。因此當應用在不同組織階層時，會因層次的區分，而給予複雜程度差異極大的資料操作授權。

### 三、簡單型企業內網路之資料發掘架構

基於前述討論，本研究提出一個簡單型的企業內網路資料發掘之整體架構，如圖 3。在簡單型中，使用者透過企業使用者端之瀏覽器，對企業內網路資料發掘伺服器發出資料發掘的相關需求，伺服器在完成資料發掘的工作後，將結果傳送回使用者後端之瀏覽器，而其自 Layer0 至 Higher Layer 的資料一般化過程，解釋如下：

(1) Layer0：在簡單型中，所有企業相關文件都以被集中於企業內網路的伺服器

上。Layer0 的資料包含兩個部份：(a) 企業內網路中伺服器中的原始網頁資料、鏈結資料以及相關的身分與授權資料；(b) 網際網路中所有的網站中的原始網頁資料、鏈結資料。

(2) 由 Layer0 至 Layer1：對於 Layer0 第一個部份的資料，以防火牆之內為範圍，使用一般資源蒐集與整理者的機制(例如主動搜尋網頁資源的搜尋引擎或是被動等待網頁創作者註冊登記的目錄服務)，將企業內

網路上目標網頁的原始資料之全部或部份由其資料所在網站上傳至伺服器的儲存空間中，然後再建立索引，進行初步之結構化。值得注意的，在此過程中，對於各份網頁所包含的相關的身分與授權資料仍須加以保存。而第二個部份的資料，則比較適宜由具有特別篩選功能的主動搜尋引擎，將網際網路目標網頁的原始資料之全部或部份上傳處理。

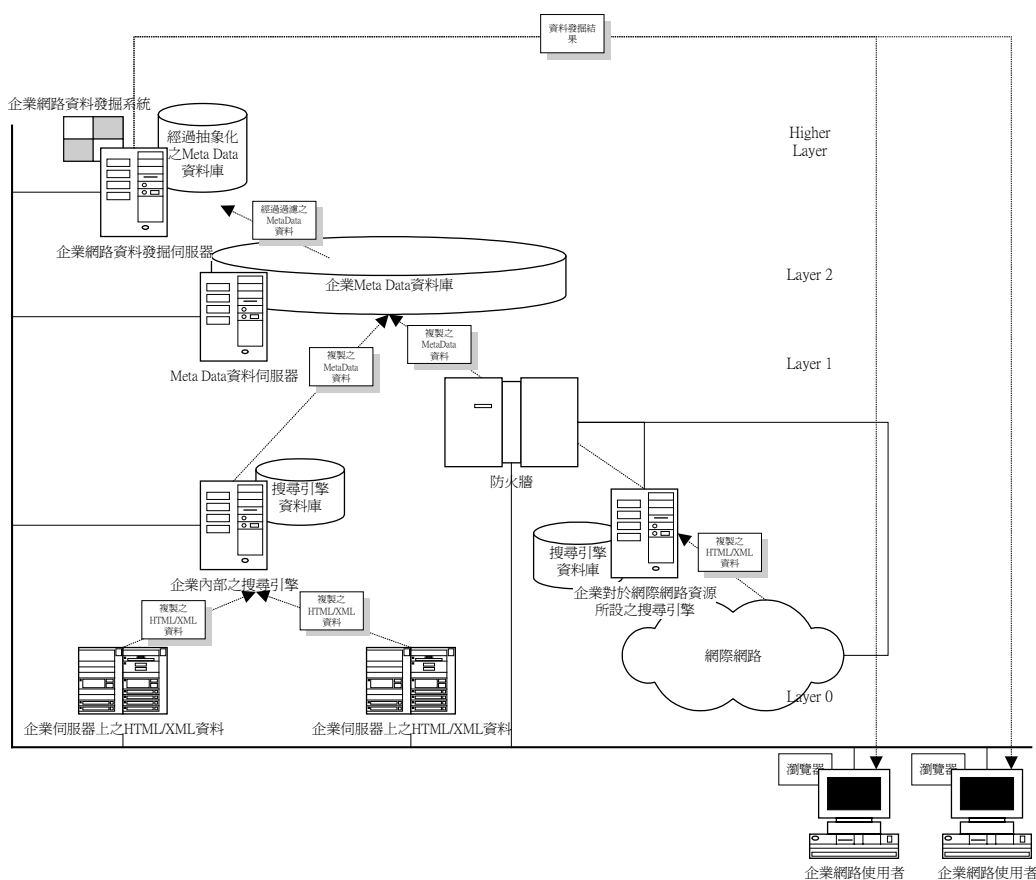


圖 3 簡單型的企業內網路資料發掘之整體架構設計

(3) Layer1：企業內網路相關資源蒐集與

整理伺服器中所保存的資料庫，包含 (a)

由前述 Layer0 的兩個部份所上傳與初步結構化的相關網頁資料、(b) 使用者對該資料查詢結果，加以儲存的衍生資料。

(4) Layer1 至 Layer2：企業內網路中各網路資源蒐集與整理者將其資料庫中具有結構化與描述性特質的 Meta Data 的網頁描述分享出來，並進行篩選與處理語意的動作，以組成一個企業 Meta Data 資源資料庫。

(5) Layer2：為邏輯上一個全企業 Meta Data 資料庫，存放所有企業內網路及可作參考的網際網路上具有結構化與描述性特質的 Meta Data 網頁描述，在實體上可能是多個分散的資料庫。這是資料發掘之主要目標資料。

(6) Higher Layer：由 Layer2 的 Meta Data 資料庫依照 Han 的概念樹學習法進行資料發掘所得之一般抽象化後之資料，此部份仍以資料庫格式儲存，唯實際已經具備知識庫的意涵。可提供進一步之資料發掘，或可供

智慧型查詢之用。

所以在簡單型的架構中，使用者端是使用單純的網際網路瀏覽器作為接收企業內網路伺服器相關服務與資訊，且被使用之瀏覽器必須與企業內網路所應用的其他技術環境相配合。然而例如現今微軟的網際網路瀏覽器 (Internet Explorer) 和網景的網際網路瀏覽器 (Netscape Communicator) 在 JAVA、DHTML 等的支援上就並不一致，所以如果企業使用 JAVA 之 JDK 1.2 所開發的 Applet 以完成資料發掘的工作，就必須採用網景的瀏覽器；相對的，如果企業使用 DHTML 所開發的網頁以完成資料發掘的工作，就必須採用微軟的瀏覽器。因此在簡單型中，資料發掘的工作在分工上，絕大部份的工作都是由伺服器端之資料發掘系統所完成，使用者端的電腦只負責結果的顯示，如圖 4 所示。

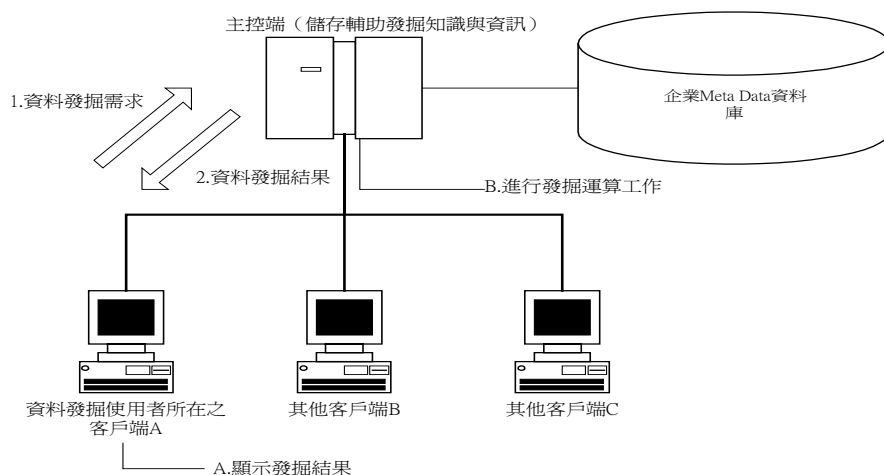


圖 4 簡單型的企業內網路資料發掘之工作分工



#### 四、複雜型企業內網路下之資料發掘架構

以簡單型架構為基礎，參考企業內網路下之資料發掘的相關思考構面，本研究提

出一個複雜型的企業內網路資料發掘之整體架構，如圖 5 所示。

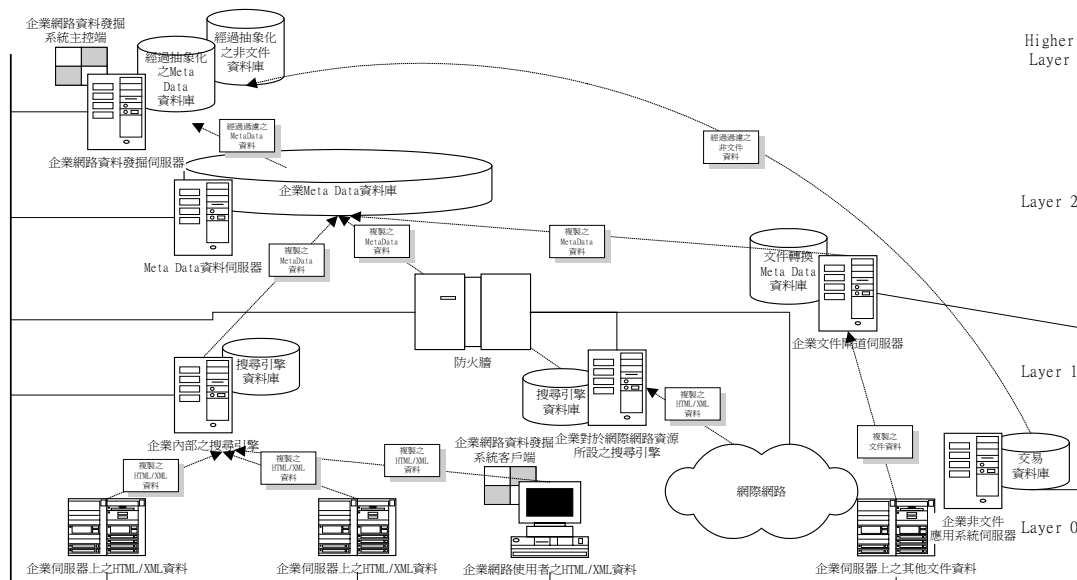


圖 5 複雜型的企業內網路資料發掘之整體架構

在複雜型的架構中，有較豐富的資料來源與較複雜的處理機制。Layer0 的資料除了有簡單型的資料外，還包括（1）在使用者端所儲存之以超文字標示語言（HTML）或是以可延伸式標示語言（XML）所構成的網頁文件的內容與相關鏈結；（2）與以上網際網路格式無關之形式在其他文件伺服器所存在之企業文件，例如早期以特別格式建置之相關群組軟體文件伺服器（如 IBM Flowmark）中的相關企業文件；（3）其他主要來自企業交易系統，以非文件方式存在之結構化資料庫。所以，由 Layer0 至 Layer1 過程中，對於其他文件伺

服器上與企業內網路或網際網路格式無關的形式所存在之企業文件，必須另行開發特定的對應技術，將其複本傳送到相關之文件通道伺服器。而其 Layer1、2 及其上的 Layers 也均包含這些透過相關通道器轉換之 Meta Data 文件描述。

另外，使用者端與伺服器端所扮演的角色分工在複雜型與簡單型兩種資料發掘架構下有很大的不同。在複雜型下，客戶端並不只是收發相關需求與結果之瀏覽器，而將包含實際負責執行企業內網路資料的發掘工作的應用系統，伺服器端並不包含完整的資料發掘系統，改而扮演供應與協調者的

角色，負責所有從儲存、管理、協調與供給相關企業整體資料發掘資源的相關工作。資料發掘的工作在分工上的可能配置也較為多樣化。假設資料發掘之主要目標資料庫（即前述 Layer2 與 Higher Layer 之資料）實體上分散在各地，而其輔助知識與資訊（資料字典、概念樹、企業規則等）有可能是集中於主控端的「**知識管理集權**」的方式，也有可能分散於各地之「**知識管理分權**」的方式，這就形成圖 6、圖 7 兩種分工的作法。當然，事實上，複雜型在分工上可能的方式還有很多，在此只是舉出兩種參考方式以說明其多樣性。

在圖 6 中，假定欲從事資料發掘工作的使用者在客戶端甲，其分工方式簡略說明如下：

（1）在集權型的分工方式中，所有資料發掘輔助知識與資訊（企業規則、資料字典、概念樹等）雖平時各客戶端甲、乙、丙各保存有自己私人的一份，但會隨時利用上傳備份機制（資料流 0）備份於主控端中。

（2）所有資料發掘的工作由客戶端甲自行運算（工作 A）與顯示（工作 B）。

（3）當客戶端私人輔助資料即足以進行資料發掘時，客戶端只向主控端要求欲發掘 Meta Data 目標（資料流 1-1），由主控端自企業 Meta Data 資料庫取回（資料流 2 與 3）。傳達回客戶端甲（資料流 4-1），客戶端甲就可逕自進行資料發掘的工作。

（4）如果客戶端私人輔助資料不足以進行資料發掘時，客戶端會在要求 Meta Data 目標同時向主控端提出輔助資料需求（資料流 1-2），由於主控端有所有輔助資訊的備份，所以可以滿足客戶端甲需求（資料流 4-2），讓客戶端甲完成資料發掘的工作。

在圖 7 中，同樣假定欲從事資料發掘工作的使用者在客戶端甲，其分工方式簡略說明如下：

（1）在分權型的分工方式中，所有資料發掘輔助知識與資訊（企業規則、資料字典、概念樹等）完全由各客戶端甲、乙、丙與主控端保存有自己的一份，不會隨時利用上傳備份機制備份於主控端中。

（2）客戶端甲向主控端提出資料發掘工作需求（資料流 1），由主控端自企業 Meta Data 資料庫取回（資料流 2 與 3）。將分配工作（工作 C）傳給各客戶端（資料流 4），各客戶端分別協同完成部份資料發掘的工作，（主控端與各客戶端甲、乙、丙之工作 A）將部份運算結果傳回主控端（資料流 5）。

（3）主控端彙整所有運算結果（工作 D），將完整的結果（資料流 5）傳回發出資料發掘需求之客戶端，再由客戶端顯示（工作 B）。

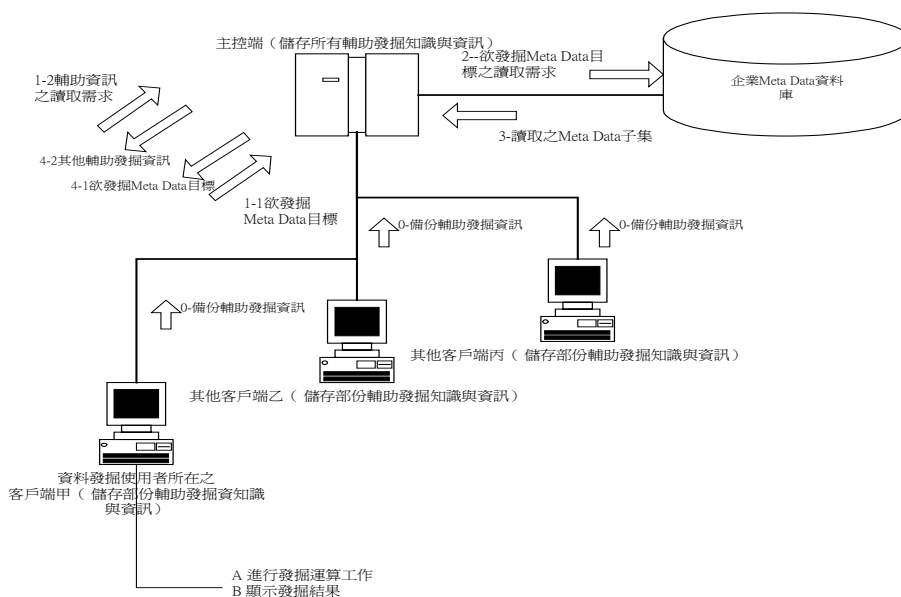


圖 6 複雜型的企業內網路資料發掘之知識管理集權型的分工方式

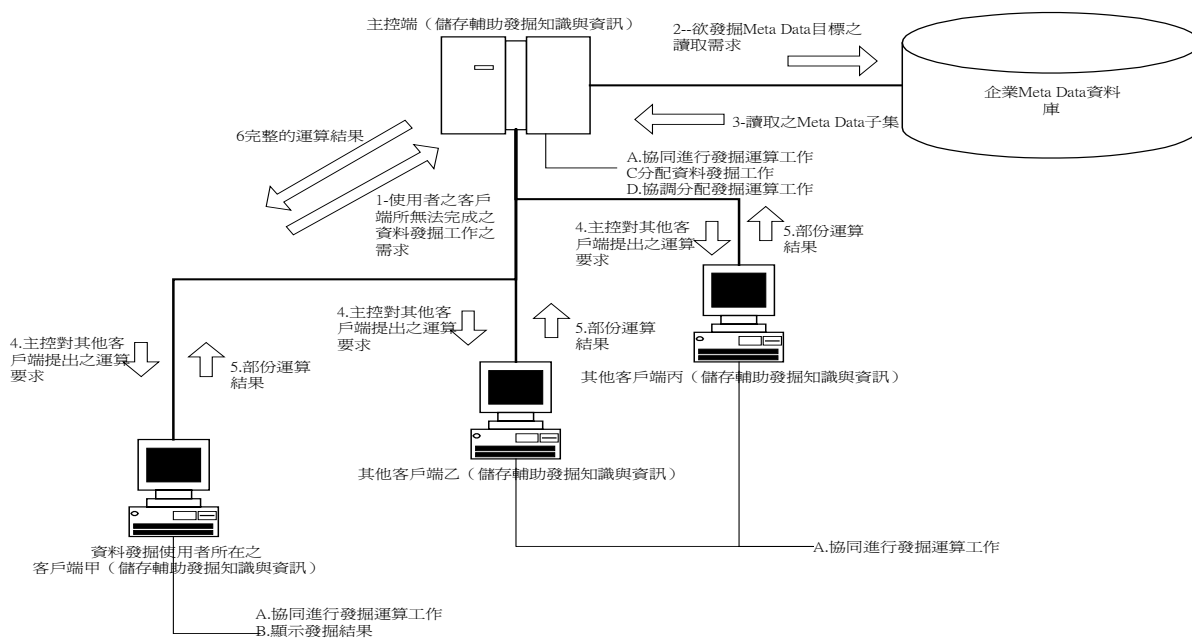


圖 7 複雜型的企業內網路資料發掘之知識管理分權型的分工方式

其實在這兩種分工方式中隱含了一些組織面的意涵。例如在集權型的分工方式中，雖然企業內網路的個別客戶端在儲存資料發掘輔助知識與資訊的內容可以隨其應用與授權而有很大的不同<sup>2</sup>，但是由於在資料發掘的過程中，單一客戶端有可能必須接收其他備份於主控端輔助資訊，並予以解讀，因此其格式與用以讀取的應用軟體必須較為統一。在分權型的分工方式中，由於各客戶端所傳回主控端的是運算後的結果，因此各客戶端不但可自主地管理資料發掘輔助知識與資訊的內容，其格式與用以讀取的應用軟體也可以不一致。因此，在分權型的分工方式中，企業中各區域享有比更集權型的分工方式更高的自主性。

## 五、以知識管理的角度檢視企業內網路下之資料發掘

如前所述，常見企業知識的取得可分為企業內部發展、由企業外部協助內部發展、企業由市場採購、企業和相關廠商合作、企業經由購併取得等五種方式。經由本研究之企業內網路資料發掘所得之以規則方式表現之企業知識，究竟是屬於以上哪一種知識取得的方式，是值得我們思考的首要

<sup>2</sup> 如同樣是收入高低的概念樹，在客戶端甲中反應出台灣的物價水準，在月收入 1000 美元以下是低所得，1000 美元到 2000 美元之間是中等所得，而 2000 美元以上是高所得。在客戶端乙中反應日本的物價水準，在月收入 2000 美元以下是低所得，2000 美元到 4000 美元之間是中等所得，而 4000 美元以上是高所得。

問題。

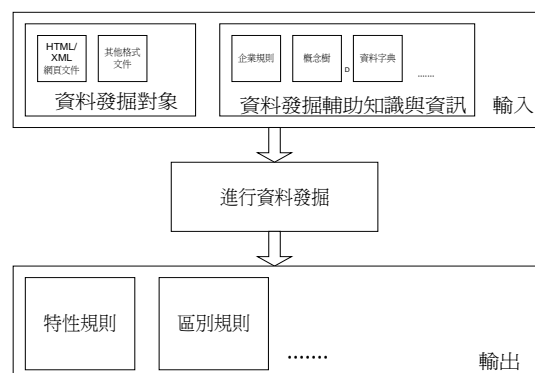


圖 8 企業內網路下之資料發掘之輸入、處理、輸出模型

若單就進行本研究之企業內網路下之資料發掘工作本身，不論是在複雜型與簡單型架構下，資料發掘工作都是在企業內部，由企業內部人員所進行。所以應是屬於內部發展的知識取得方式。但若參考圖 8 之輸入、處理、輸出模型，再進一步思考此一問題時，我們發現由於作為輸入的發掘對象資料並不只限於企業內部所產生的網頁文件，也可能包含自網際網路所蒐集而來的網頁文件、甚至是自其他企業取得之文件。而作為輔助資料發掘工作中的企業規則與概念樹等知識，也可能包含並非由企業內部所產生的部份，例如由顧問公司所提供的一些特殊觀點與角度的概念樹等。所以經由企業內網路下的資料發掘，兼具「企業內部發展」與「由企業外部協助內部發展」的知識取得方式。

其次就**知識的整理**來看，雖然經由企業內網路下資料發掘所得之各種規則，理論上已屬於格式相當確定的顯性知識，可作為知識轉換與學習的目標。但是以企業內網路下之各種網頁與非網頁文件蕪雜的情形來看，資料發掘所得之規則在可留下保存前，仍必須進行去除無趣規則的整理。所謂無趣的規則是指：發掘所得的規則是可預期的或是符合已知的知識、或是由無趣的屬性組成、或發掘所得的規則是多餘的。至於去除無趣的規則的工作，除了手動去除的方法外，也可以利用統計學上相關檢定來評估規則，或定義期望值函數，若該規則產生的值不大於期望值，則予刪除（可參考如 Piatetsky-Shaprio, 1991；Siberschatz and Tuzhilin, 1995）。

再就**知識的轉移與學習**來看，如果以知識轉移 = 傳達 + 吸收 + 利用的模式來看，企業內網路下資料發掘本身就可以視為一種知識轉移的過程。它包含了文獻探討中，四個知識轉換模式其中的三個：

(1) 內隱至外顯（外化）：在資料發掘的架構中，其準備工作包含兩個內隱至外化的部份，其一在於將非結構化的網頁利用內顯知識進行 Meta Data 的建立與定義，轉換成爲較具結構化的資料。其二則在於企業成員將其所具有的內隱知識明白表達爲資料發掘的外顯輔助知識（企業規則、概念樹等）。

(2) 外顯至外顯（結合）：在資料發掘的架構中，其準備工作有時需要結合不同選擇的資料發掘對象、輔助知識（企業規則、概念

樹等）等，形成另一些同屬外顯形態的輔助知識。而資料發掘的核心演算法則結合不同的目標資料、輔助知識，轉化爲同屬外顯形態的規則（區別規則、關連規則等）

(3) 外顯至內隱（內化）：在進行資料發掘後，以外顯形態表現的規則，在經過上述去除無趣的規則等整理後，可利用企業內網路再傳播，使企業內成員透過閱讀這些規則，內化成該企業整體或成員個人有價值的資產。

由於企業內網路下資料發掘架構爲一組支援資料庫或知識庫且在相關資源的授權與管理上相當完整的資訊技術，因此可將經由整理過的發掘所得之規則直接儲存在企業內網路的知識庫中，再建立相關資料發掘結果的查詢與管理系統，來蓄積組織長期的知識基礎。並且在資料發掘過程中，也可以透過利用前人發掘結果所得規則作爲新發掘過程之輔助知識，達成知識在組織中改進與傳播的作用。

最後要談到**知識分享**，當組織成員向他人學習知識時，就是在分享他人的知識，所以引用他人所建立之企業規則、概念樹等知識在企業內網路下來進行資料發掘，實際上也是組織成員間一種知識學習行動。而資料發掘過程也可以視作知識接收者重建與擴展知識的行爲。在這過程中，資料發掘之「知識需求者」其實需依賴擁有顯性知識的規則或概念樹之「知識擁有者」。

若假設整個企業的顯性知識爲其全體成員個別可以接觸到的顯性知識之聯集。對

於組織中個別成員來說，在整個企業的顯性知識聯集中，只有符合其經由溝通網路、友誼網路（人際關係）交互作用影響之**工作流程網路**（各種角色有不同的階層以及地區上的工作需求）授權的部份才為其可知與可取用，對這我們可稱為「組織內個人的顯性知識視野」，如圖 9 所示。

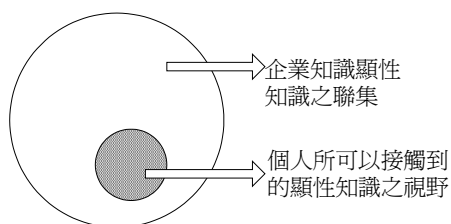


圖 9 個人對於組織中蓄積之顯性知識的視野

引申上述的概念，由於簡單型的資料發掘的分工，絕大部份的工作都是由主控器端之資料發掘系統所完成，使用者端的電腦只負責結果的顯示。而企業對於主控端各種資料發掘的目標資料文件與輔助資源的授權與限制，是依照組織中每個人在工作流程網路不同的角色扮演。所以在簡單型下，每個人在資料發掘過程中知識的視野幾乎完全來自於工作流程網路的角色，如圖 10。

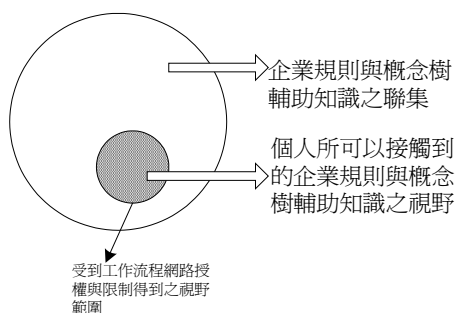


圖 10 簡單型企業內網路架構下組織中個人對

#### 於輔助發掘知識的視野

而在複雜型下，若為集權型的分工方式中，對所有資料發掘輔助知識與資訊（企業規則、資料字典、概念樹等），理論上各客戶端所在的事業分支單位隨時會受到總管理處的監督與控制，利用上傳備份機制將之備份於主控端中。在進行資料發掘的過程時，輔助知識與資訊皆由主控端負責供給，這種分工方式其實意味對組織內的顯性知識進行集中管理。此時，組織中每個人對於資料發掘輔助知識的視野，在最理想的情形下，應該最多受到知識上傳更新延誤的影響而比工作流程網路所授權的略小。但是在實務上，由於知識本身在內隱與外顯間可以互相轉換，而且不易於像交易資料可訂定追蹤與稽核的標準。因此以下兩種因素的干擾會讓組織中發生某些知識呈現時顯時隱的狀態，而使每個人對於資料發掘輔助知識的視野更狹隘：（1）個人（或部門）對於集中式管理權力的抗拒；（2）個人（或部門）間溝通網路與友誼網路中的人際關係瓜葛。也就是說，若個人（或部門）可能會因為基於一些抗拒集中管理的心態、或溝通網路與友誼網路中非理性的情緒，而將資料發掘輔助知識與資訊平時保存在私人電腦空間中（或以隱性知識的方式保存在個人的記憶中），當其自己需要使用時才臨時加入系統之中，用完即將之刪除。這將造成個人所使用的輔助知識範圍可能會比理論上工作流程網路的角色限制的更狹隘。由於這種妨礙知識分享的行為將對全體造成同樣的傷害，因此干擾

發生的機率（如圖 11 中的粗實線）主要與總管理處權力的大小，以及總管理處的相關人員與各事業分支單位人員在溝通上或人際關係的好壞成反比。當然各事業分支單位人員間溝通網路與友誼網路中的特定人際關係也會有所影響，不過基於妨礙知識分享的行為將對全體造成同樣的傷害，所以除非其對於分享的其他某事業分支單位人員相當確定，否則其干擾發生的機率（如圖 11 中的細虛線）較小。

在分權式的分工中，所有資料發掘輔助資訊完全存於各客戶端所在的事業分支單位，不會上傳給主控端，而進行資料發掘的過程時，主控端只進行相關協調，不負責全部輔助知識與資訊的供給，這種分工方式

其實意味對於組織內的顯性知識進行分散管理。若總管理處進一步允許個人（部門）可以設定自己電腦對於同儕的授權與限制，就更易發生針對特定對象進行妨礙知識分享的行為，且不必擔心對其他多數群體造成影響。若總管理處不允許個人（部門）設定自己電腦對於同儕的授權與限制，也可能發生將電腦惡意關機以規避分工的困擾。所以組織中溝通網路與友誼網路的人際間（或部門間）關係將形成對工作流程網路的授權更大的干擾。此時，若無一套思慮周全的知識管理機制，個人（或部門）對於輔助資料發掘相關概念樹與改寫規則的顯性知識視野，會比集權型的分工方式更顯得狹隘與錯亂。如圖 12 中個人間的粗實線所示之影響。

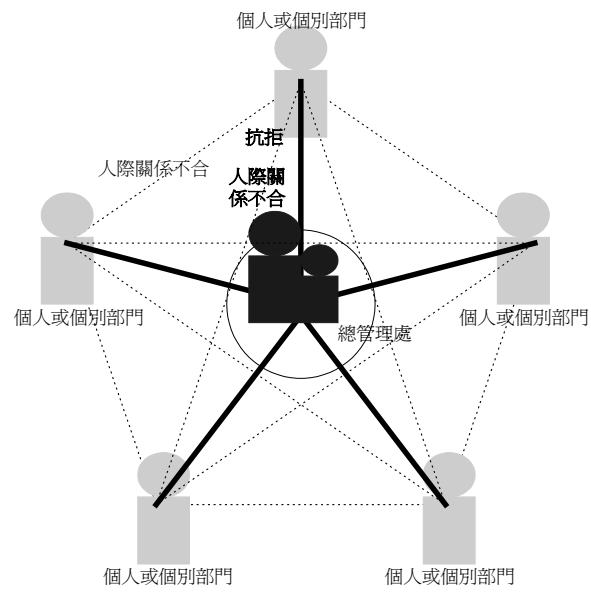


圖 11 複雜型企業內網路架構下集權型的分工方式之知識視野所受影響示意圖

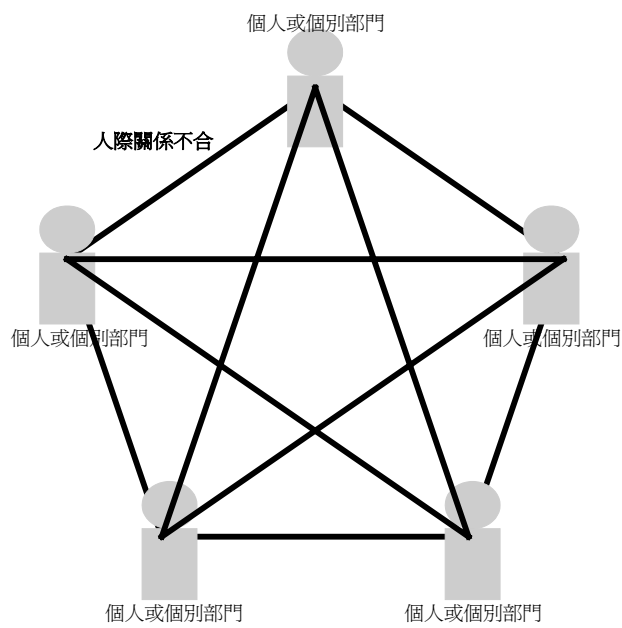


圖 12 複雜型企業內網路架構下分權型的分工方式之知識視野所受影響示意圖



最後有兩點是值得特別說明的，依本研究的架構，所有資料發掘的目標自Layer0至Higher Layer都保存在企業內網路資料庫系統中，受到資料庫管理系統相關的授權規範與管理，這些目標資料授權理論上並無問題。本文上述所討論的顯性知識視野問題，則是針對研究架構中輔助資料發掘的概念樹與改寫規則的知識授權問題。其次，則是若以知識分享的角度而言，在資料發掘的過程中，一些與發掘方法 (Methodology) 相關的隱性知識更可能衍生知識授權與視野的問題，這有待未來進一步的討論。

## 六、企業內網路下的語意表頭

基於本研究以往對網際網路語意表頭 (Semantic Headers) 的研究成果 (楊亨利、金士俊，民 90 年)，再考量企業內網路的特性，進一步分別提出簡單型與複雜型之語意表頭項目，兩者均包含下列 19 項：。資料

來源 (Source)、主題 (Subject)、標題 (Title)、日期 (Date)、版本 (Version)、相關人員 (RespAgent)、分類 (Classification)、類型 (Genre)、確認字串 (Identifier)、與其他文件的相關性 (Relation)、語言 (Language)、字元集 (CharSet)、關鍵字 (KeyWords)、系統需求 (SysReq)、涵蓋範圍 (Coverage)、應用擴充領域 (Applications)、開發工具 (Generator)、重要性 (Importance)、授權 (Authorization)。其中「應用擴充領域」是為容許不同應用領域彈性擴充之用。而考量了前述簡單型與複雜型差異的幾個維度，雖兩者均有相同的 19 項，但其表達之精細程度並不同。如表 1。

表 1 簡單型與複雜型 Meta Data 之例

	簡單型	複雜型
資料來源	<META NAME = "Source" CONTENT = "Intranet">	<META NAME = "Source " CONTENT = "SrcDomain =NIC_MAC_SERIAL, CovSchema =XXXXXXXXXX, CovValue =AABBCCDDEE">
註釋	這份文件是企業網路內的文件	此份文件的資料來源是來自網路卡號碼以XXXXXXXXXXXX格式表示時之 AABBCCDDEE 卡號的電腦(NIC_MAC_SERIAL 是指電腦上的網路卡出廠全球識別編號，類似的例子還有硬別編號等，常被軟體公司用來鎖定電腦以防範侵害智慧財產權之預防措施)。
重要性	<META NAME = "Importance " CONTENT = "TSchema =mm/dd/yyyy, Init_Time=05/02/2000, End_Time=05/03/2000, Target=Kiso ,Action=Read, Degree=High">	<META NAME = "Importance " CONTENT = "TSchema =mm/dd/yyyy, Init_Time=05/02/2000, End_Time=05/03/2000, Target=Kiso , Level=Management_Control Action=Read, Degree=High">
註釋	此份文件在 2000 年五月二日到 2000 年五月三日之間，對於 Kiso 具有閱讀上的高度重要性	此份文件在 2000 年五月二日到 2000 年五月三日之間，對於 Kiso 在 <b>管理控制的應用層次</b> 上具有閱讀上的高度重要性

## 七、企業網路內資料發掘演算法

我們認為不論是在簡單型或是複雜型企業網路的架構下，超文字結構標準的形成皆是必然的趨勢，因此假設企業內超文字文件提供者都必須填註本研究所定義之 Meta Data，而企業中搜尋引擎與目錄服務的伺服器自然必須配合所蒐集的 HTML/XML 的 Meta Data 分享出來，故可假設 Layer2 的企業 Meta Data 資料庫均已經建立。所以，將研究焦點先集中於之各個架構下發掘模組的發展。本研究分別對簡單型、複雜型集權式分工、分權式分工三者提出其演算法。由於篇幅所限，在此僅對大的流程略加描述。圖 13 即是簡單型之發掘模組演算法流程，整個流程可以區分為四個不同的階段：

- 階段 A：依照驗證使用者所輸入的身分、決定授權範圍，挑選合適的資料發掘資源(包括合適與符合授權的資料來源、知識庫以及概念樹)。
- 階段 B：劃定發掘資料範圍。從使用者對於應用領域的挑選，使用者對於一般網際網路 Meta Data 表頭屬性與應用領域屬性的挑選與條件篩選，到將專門名詞的置換，並把所有的篩選與挑選工作以企業改寫法則對於原始查詢作最佳化，一直到起始查詢形成為止。

- 階段 C：運用修正之概念樹演算法，進行資料發掘。從使用者對於資料發掘作相關設定，根據不同的資料發掘需求與概念樹的組成，對於起始表格進行相關處理(屬性轉換，屬性分解，屬性合併等)形成發掘表格，然後不斷循環對於發掘表格進行抽象化與相關處理的修正工作以發掘資料。
- 階段 D：發掘結果的呈現。最後對於發掘結果進行轉譯與解讀的工作，將結果呈現給使用者。

複雜型集權式演算法流程與簡單型，主要的差別是在階段 A。在複雜型集權式下，階段 A 之相關驗證使用者身分與決定授權範圍、挑選發掘資源等工作會先在客戶端進行。當客戶端儲存之相關資源實有不足時，才由主控端依據使用者的身分與授權資料的驗證，由其歸屬之群組，列出各種可供選擇之資源清單供其挑選，待使用者分別在客戶端與主控端。

挑選資源完畢，系統必須對於二者進行整合。其他階段演算法流程中之其他工作則將均在客戶端完成。而複雜型分權式演算法流程則變得十分複雜。從階段 A 開始一直到階段D，工作分別會在主控端與客戶端以及其他客戶端進行。且需應用平行處理技術分別在客戶端、主控端與其他支援之客戶端作專門名詞的置換、根據不同的發掘需求與概念樹對於起始表格進行相關處理、抽象化等工作。

### 肆、系統雛形實作

#### 一、雛形系統環境

根據前述架構，本研究假設存在兩家相同產業<sup>3</sup>(如英語教學補習)的機構，一家為使用簡單型企業內網路架構的小型本土業者，另一家則為使用複雜型的企業內網路架構的大型國際連鎖業者，分別建立其模擬的企業內網路環境與其對應之Layer 2 Meta Data 資源資料庫，然後建立雛形系統 (Prototype)，以驗證本研究之可行性。

本雛形系統所使用的軟體發展環境較為複雜，主要分為三個部份：(1) 企業內網路環境：以 Windows98 平台搭載簡易的 Winproxy 防火牆軟體在單機上進行企業內網路環境的模擬；(2) 企業全

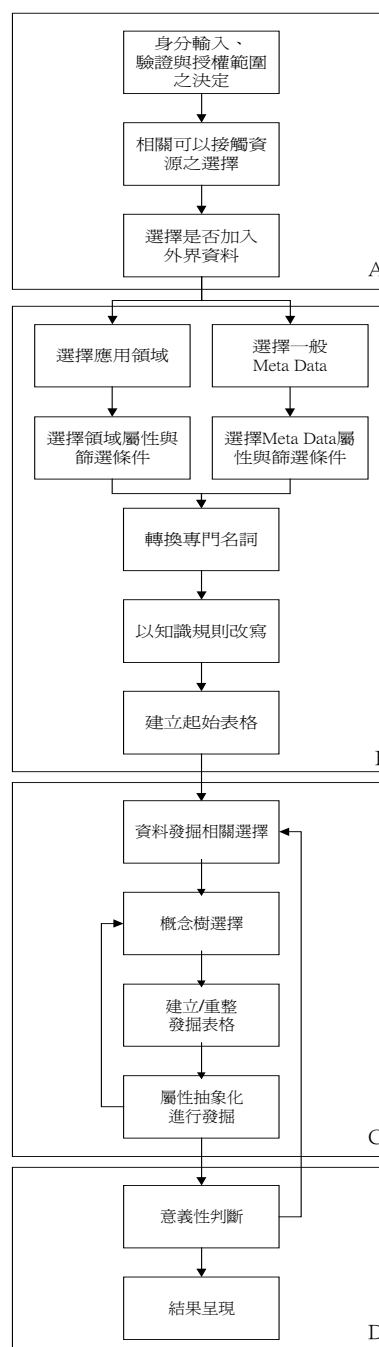


圖 13 簡單型企業網路架構下發掘模組演算法流程圖

<sup>3</sup> 為求簡化，本雛形並未假設文件之應用擴充領域項目。

球資訊網伺服器：包括微軟的 IIS (Internet Information Server)、ASP (Active Server Page) 以及嵌入的 Java Applet 元件；(3) 資料庫伺服器：主要為 Microsoft 公司的 SQL Server。而在 IIS 與 SQL Server 之間以 ODBC (Open DataBase Connection) 相連，

JAVA Applet 則藉由 JAVA 所提供的 RMI (Remote Method Invocation) 機制，透過自行撰寫的 RMI-JDBC 伺服器以 JDBC-ODBC 橋接器與 JDBC 和 SQL Server 相連。如圖 14 所示。

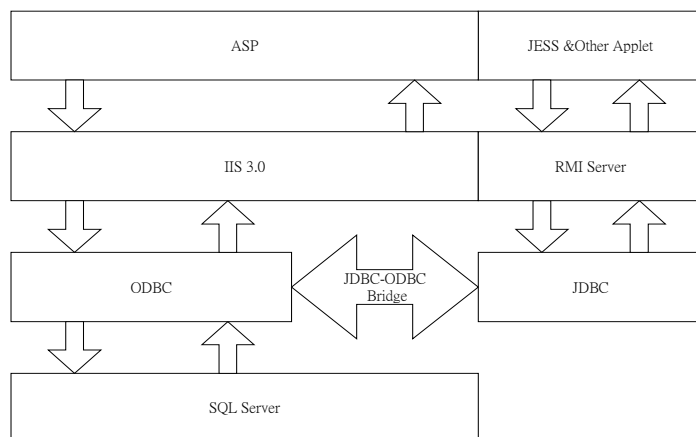


圖 14 系統軟體環境架構圖

如以使用者的角度來觀察，在本雛形中，使用者界面係以 ASP 程式語言，於 IE 瀏覽器中建立視窗的圖形化界面，供使用者點選並輸入需求，提供使用者方便的界面操作，將結果以規則及表格的方式呈現給使用者。

在改寫規則部份，本研究利用國外發展已有一定時日的 JESS (Java Expert System Shell)，加強其與全球資訊網資料庫方面的連結，以獨立的 Java Applet 嵌入 ASP 中，並在 SQL Server 中建立知識庫處理，以進行對於所有改寫規則的處理。所有改寫規則均使用 CLIPS 語言，以「條件」與「對

應行動」方式呈現，存在於 SQL Server 資料庫表格中，可由有權管理者以一般資料庫存取方式加以修改。

## 二、雛形系統相關資料庫結構

本雛形的資料庫群可以區分為四個部份，如圖 15 中之 1、2、3、4 塊。目前實作了簡單型與複雜型的集權式分工兩者的資料發掘。由於複雜型的資料來源種類很多，又有地區性、部門觀點差異，故其相關的資料庫群之第二、三、四部分均有多組資料的現象。

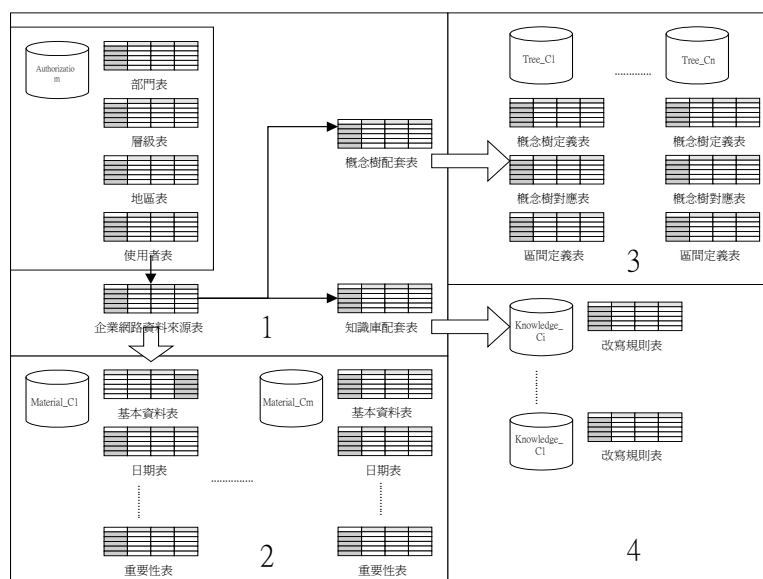


圖 15 資料庫群組成結構示意圖

(1) 第一部份：每一企業需要一組用以對於使用者登入進行身分確認與授權相關資源使用的授權資料庫，包含部門表、層級表、地區表、使用者表（某使用者所擁有的部門、層級、地區授權）、原始 Meta Data 資料授權表（某組 Meta Data 開放給的部門、層級、地區授權）、概念樹配套表（某組 Meta Data 所對應的概念樹組）與知識庫配套表（某組 Meta Data 所對應的改寫規則組）七者。

(2) 第二部份：對應於第一部份之原始資料授權表，本雛形建立五組作為資料發掘對象的資料項目資料庫（複雜型集權式分工方式四組，簡單型一組），一組內共含有十六

個 Meta Data 表格<sup>4</sup>。彼此之間以網址作為外鍵互相連結。而其中的內容，由於目前填入完整 Meta Data 的資料網頁數量有限且蒐集不易，因此使用自行發展的 EXCEL 巨集作為原始 Meta Data 資料項目資料，各產生一百筆左右之資料，作為雛形資料發掘的對象。

(3) 第三部份：對應於第一部份之概念樹配套表，本雛形建立六組用以輔助資料發掘進行抽象化之概念樹資料庫（複雜型集權式分工方式五組，簡單型一組），每一組中包含概念樹定義表（定義出每棵概念樹之層級

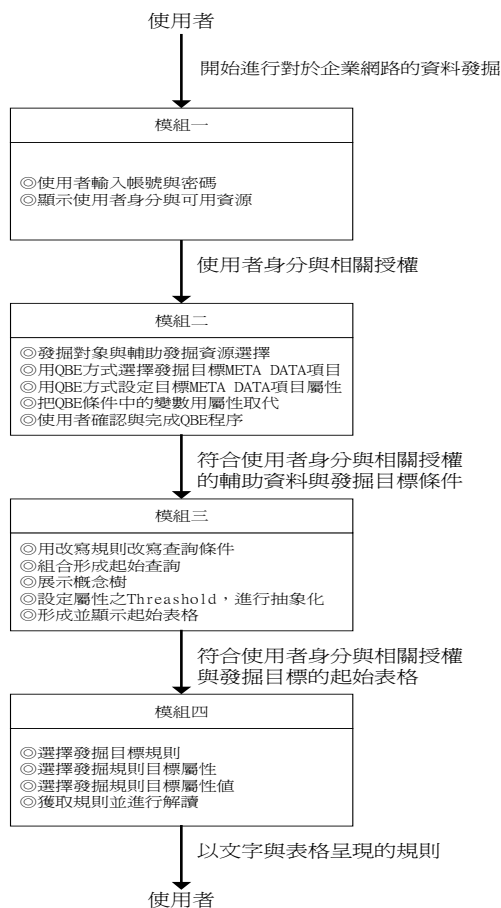
<sup>4</sup> 前述建議有十九個 Meta Data 項目，本雛形在此未建立擴充領域項目，並將「標題」、「語言」、「字元集」三者合併於「基本資料表」中，其餘十五項分別與此處表格對應。

結構)、概念樹與屬性對照表(哪些Meta Data表中之屬性有對應之概念樹)、概念樹區間對照表(對於屬數值型資料之區間定義)三者<sup>5</sup>。

(4) 第四部份:對應於第一部份之知識庫配套表,本雛形建立六組用以改寫使用者查詢的一般知識與企業知識的改寫規則資料庫(複雜型集權式分工方式五組,簡單型一組),每一組包含一個改寫規則表表格<sup>6</sup>。各組改寫規則均存有一般常識型的改寫規則八十三條、企業內網路的改寫規則十條。

### 三、系統雛形模組與運作流程

複雜型集權式分工的雛形可以區分成四大模組,如圖16所示。由於我們對部門、層級與地區的ID編碼均個別採質數<sup>7</sup>,所以,在模組一中,當使用者登入時,在通過ID和密碼的檢查後,系統便根據使用者表,取得該使用者相關的授權因子(某合成數)與部門表、層級表與地區表中的ID比對,若不互質,則代表該使用者享有該部門、層級或地區的授權。同樣的,對Meta Data資料來源組之ID編碼也採質數。模組二讓使用者對於所選取之Meta Data組內所有表格的屬



性,用QBE (Query By Example) 的方式設定篩選條件。但因未實作專有名詞辭典,所以所下的條件內,不可包含資料庫中沒有涵蓋的值。在模組三中,對於前面的篩選條件,運用專家系統對於企業內網路之的知識,做適當的過濾與改寫以形成起始查詢,並在設定抽象化之屬性值個數上限(Threshold)後,初步抽象化的運算得到起始表格。對於概念樹,是採取整組選取的方式,並不允許跨組選擇概念樹。在模組四中,經過與Han類似的概念樹演算法等一連串步驟,從起始表格中找出特性或區別規則,並加以展示。另外,透過本雛形「上一

<sup>5</sup> 由於篇幅所限,有關此三者之詳細Schema可參考本研究在網際網路資料發掘之研究設計(楊亨利、金士俊,民90年)。

<sup>6</sup> 有關此改寫規則表之詳細Schema也請參考本研究在網際網路資料發掘之研究設計(楊亨利、金士俊,民90年)。

<sup>7</sup> 若部門、層級與地區不很多時,我們採用的均為小質數,比較是否互質很容易。

步」、「下一步」的方式，使用者可以從同一棵概念樹相同概念層次的不同屬性值的選擇中，得到同一概念層次的多個規則，而不必像Han之演算法，即使是同概念層次的相關規則發掘，每一次資料發掘都必需重頭從起始表格開始作起。

#### 四、 雛形展示

爲了展示雛形中的一些機制，我們假設前述大型國際連鎖業機構欲從本研究所建立之虛擬企業內網路空間資料中找出「存在於企業內網路中某一特別範圍 URL 位址爲 10.10.0.0~10.10.255.255 (URL 位址爲 10.10.xxx.xxx)，包含來自企業內網路網站上的網頁文件與透過閘道器轉來的其他文件等。針對不同的創造軟體、表現在創造日期與檔案尺寸上的概念特性爲何？」；另一方面，也想了解「符合同樣條件的創造軟體間在概念上如何區分？如果引用企業內不同觀點的概念樹將會得到哪些不同的知識？」。以下即爲此模組之展示。

我們並假設 Meta Data 文件開發工具表中，所用之 Software 由於不同部門的觀點差異，有兩棵概念樹。一棵 Tree\_C5\_1 是一般使用部門的觀點，對他們而言，文件的製作軟體第一層可能分爲 Office (一般辦公室使用)、Internet (網路使用)、DTP (桌上排版) 三者，如圖 17。另一棵 Tree\_C5\_2 是文件開發部門的觀點，對他們而言，文件的製作軟體第一層可能分爲 Content\_Creator (以撰寫內容爲目的)、

Layout\_Creator (以排版爲目的)，如圖 18。

使用者先登錄使用者 ID 與密碼，並且由系統確認使用者的身份與相關授權(包括工作相關部門、工作相關層級，及其所屬地區)，以及決定其可使用的資源後，即可由其在圖 19 (爲第二模組之一畫面) 選擇資料發掘的原始資料對象。根據使用者選擇資料發掘的目標資料，系統配對輔助資料發掘之規則知識庫與概念樹群，如果有一個以上，則由使用者進行選擇。在本範例中，假設我們複選 Material\_C1 (企業內網路網站上的網頁文件) 與 Material\_C2 (透過閘道器轉來的其他文件)，對於適用於此一組合資料發掘的概念樹，由於此處只能單選不能複選，所以假設我們先選擇第一組 Tree\_C5\_1，而企業規則知識庫選 Knowledge\_C5。

使用者再在圖 20 (也是第二模組畫面之一)，使用者用QBE方式選擇發掘目標 Meta Data項目，並可對所選擇的Meta Data的表格相關屬性進行更進一步的設定。在本例中，我們選取Meta Data中開發工具之軟體屬性、日期之創造日期與日期格式屬性，以及檔案類型之尺寸屬性，並設定基本資料表中之URL。此後，系統將進入第三模組。系統先根據使用者在第二模組所選定之企業規則知識庫 (Knowledge\_C5)，將前面第二模組所有輸入的篩選條件改寫成Clips的語句，運用一般常識型的改寫規則、一般國際網路的改寫規則、以及應用領域的改寫規則的知識庫對其進行改寫規則的處理，處理

完的Clips語句，再改寫回原有的篩選條件的形式傳達出來。然後系統將形成起始SQL查詢<sup>8</sup>。如果他對SQL語法熟悉，也可以別對如圖 22 中的三部份的SQL加以修改。不過本雛形對於此部份並無文法檢查，輸入錯誤將導致無法正確進行資料發掘的相關工作。

然後在模組三中，系統會在螢幕展示其所選擇的概念樹（如圖 17），使用者可依此設定抽象化之屬性值個數上限。系統會執行前述 SQL，形成抽象化之三部份的起始表格，並將此三個部份的起始表格以網址主鍵鏈結連結（Join）成爲單一起始表格。第四模組對此起始表格進行資料發掘並得到結果（如圖 22）。如果在第四模組功能一選擇特性規則、功能二選擇 Software 爲特性規則目標組屬性、功能三選擇目標組屬性值爲 Office，可以得到之特性規則如圖 22 所示。簡略來說，此特性規則是指「在 10.10.0.0 ~ 10.10.10.255.255 的子網路中，60% 用 Office 軟體創造的文件都是很久以前創造的，尺寸都較大。」

若我們再進入模組二功能一中，同樣複選 Material\_C1 與 Material\_C2，但這次概念樹，我們選擇 Tree\_C5\_2（如圖 18），進行資料發掘。得到之特性規則簡略來說，將會是「在 10.10.0.0~10.10.10.255.255 的子

網路中，50% 用 Content\_Creator 軟體創造的文件都是很久以前創造的，尺寸都是中等的」。

經由解決此一之資料發掘命題的系統雛形展示，我們除了展現系統實作的成果外，更重要的是表達了對於相同的起始查詢目標，引用不同的概念樹，將會獲得截然不同的知識的觀念。在此範例中，我們雖假設概念樹的不同是源自不同工作部門，其實即使同一部門不同的知識專家，也會有不同的概念<sup>9</sup>。另外，因爲使用者階層與地區的差異而導致不同的概念樹也是很常見的，下一個範例就將焦點放在地區不同而導致不同的概念樹上。

第二個範例命題如下：「欲找出存在於企業網路中不同地區的文件中，針對不同的語言、表現在創造日期與檔案尺寸上的概念特性爲何？如果對不同地區並引用企業內因地制宜而有不同觀點的概念樹，而兩地區發掘目標資料恰巧雷同將會得到如何不同的結果？」我們並假設此機構有台灣與香港兩個地區分公司。所以，有兩個地區性的資料來源，分別爲 Material\_C3 與 Material\_C4。爲比較起見，我們故意讓兩者包含相同的資料。另外，其 Language 之概念樹有二，Tree\_C3、Tree\_C5 分別適用於台灣與香港。對於台灣地區分公司的人來說，台灣語（Taiwanese）是一種官方語言（Official Language），而英語（English）是

<sup>8</sup> 這裡其實有複合屬性概念樹的問題，所以系統其實會產生單一屬性、複合二屬性、複合三屬性三部份SQL，詳細可參考本研究在網際網路資料發掘之研究設計（楊亨利、金士俊，民 90 年）。

<sup>9</sup> 後者即是Han的多重概念樹的涵義。



一種外語 (Foreign Language)。但對於香港地區分公司的人來說，英語則是一種官方語言，而台灣語是一種外語。

在使用者 Eric 登入後，經由系統模組一功能二檢核出此使用者擁有台灣與香港兩個地區的身分授權。在進入模組二功能一中，先選擇 Material\_C3 (台灣企業網路網站上的網頁文件)，對於適用於此一組合資料發掘的概念樹，只有 Tree\_C3，開始進行資料發掘。經由 QBE 界面進行相關的選取與設定、若其設定抽象化屬性值個數上限為 3、選擇特性規則、目標屬性選擇 Language、目標屬性值選擇 Foreign\_Language。可以得到之特性規則，

簡略來說，是指「在台灣分公司中大部份外語的文件都是很久以前創造的，尺寸都是中等的。」

若我們再進入模組二功能一中，選擇 Material\_C4 (香港企業網路網站上的網頁文件)，對於適用於此一組合資料發掘的概念樹，只有 Tree\_C4，開始進行資料發掘。並對系統作與上述類似的選擇，可以得到之特性規則，簡略來說，將是指「在香港分公司中一半以上外語的文件都是很久以前創造的，尺寸都是小的」。可見即使不同資料來源內含相同的資料，但因一些因地制宜的概念並不一致，會發掘與解讀出不同的規則。

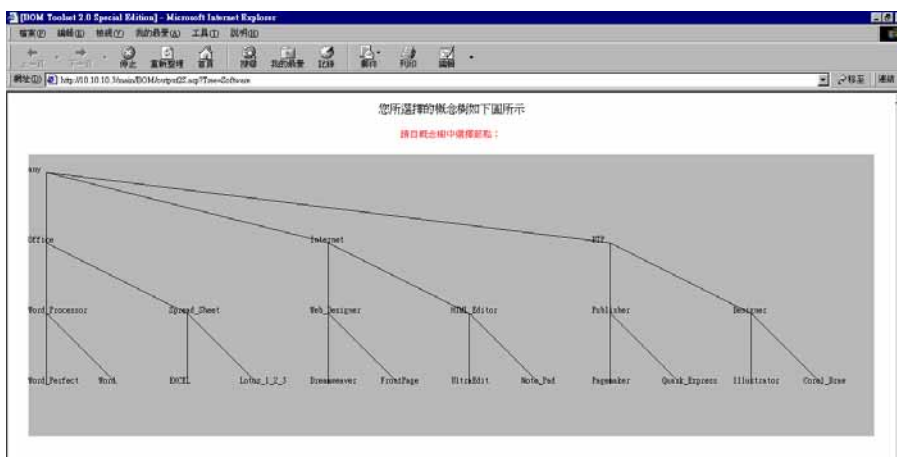


圖 17 Software 概念樹 Tree\_C5\_1

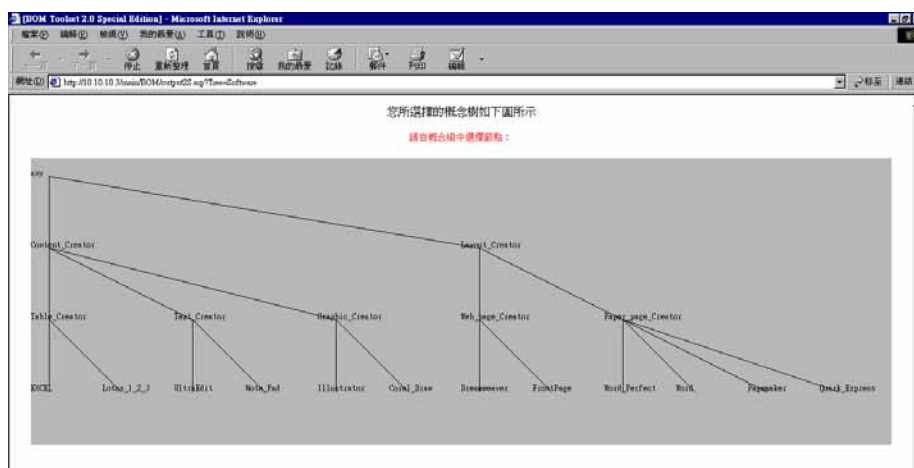


圖 18 Software 另一概念樹 Tree\_C5\_2



圖 19 雛形中第二模組之一畫面 (可選目標資料來源與輔助知識)



圖 20 雛形中第二模組之一畫面（可設定 QBE 之條件）



圖 21 雛形中第三模組之一畫面（列出三部分之 SQL）



圖 22 雛形中第四模組之一畫面（列出發掘結果）

## 伍、結論

### 一、本研究之貢獻：

#### (一)學術上的貢獻：

(1) 架構上的貢獻：本研究提出對於企業內網路的思考構面、複雜型與簡單型企業內網路資料發掘的參考架構，以及用不同知識觀點對待同一資料發掘對象的觀念。

(2) 系統實作上的貢獻：本研究的雛形有下列優點：以 RMI 技術建立 JESS 與資料庫的連結；運用 QBE 具有親和性的使用者界面，讓使用者較易下達其需求；系統模組具備獨立性、擴充性；實作因使用者授權與選擇差異而衍生的多重輔助知識觀點。所產生的雛形應屬以企業內網路的授權問題與知識視野角度，建立以企業內網路內容資料發

掘研究的先驅者之一。

#### (二)實務應用上的貢獻：

由於企業內網路已經成爲一個企業的重要資源，企業必須制定合理的決策對其加以策進與規範，經由本研究的建議，企業可依所提出之思考構面，檢視其在企業內網路上簡單與複雜程度應具備的架構。而對於企業一般使用者，可由本雛形中了解企業內網路上的相關網頁與其他文件，是依照何種知識或規則的方式存在，進一步形成其企業決策的知識來源。例如，可以參考網頁裡關於產生者之軟體的規則與知識，作爲採購與教育訓練的相關規劃。

### 二、本研究雛形之限制：

本雛形目前有下列限制：在授權系統

的設計上，目前無法容許相同人員在不同部門或不同地區間擁有不同階層的較複雜授權模式；輔助資源（如概念樹組）是伴隨 Meta Data 資料來源配置；不允許使用者跨組選擇概念樹；改寫規則需人工預先設計；缺乏專業辭典的輔助；未具備 QBE 全部之功能（如無法下「不存在」某些條件）；在可由專業的使用者自由修改的 SQL 文字框內，欠缺文法的相關編譯器（Compiler）檢查；規則的展現仍不夠淺顯易懂等。

### 三、建議

未來研究方向在學術上可對知識授權、知識視野與企業內多重概念樹的進一步討論；並可繼續定義更周全的企業內網路 Meta Data，及找尋其概念樹。在雛形實作上，可試圖突破上述限制，並發展更能滿足實務上效率需求的演算法。

### 參考文獻

- 【1】 江姿慧，使用者參與虛擬社群之行為研究，台灣科技大學資訊管理系碩士論文，民國 89 年。
- 【2】 周立平，從資料庫中發現法則：用學生修課的資料作分析，淡江大學資訊工程研究所碩士論文，民國 84 年。
- 【3】 陳年興、陳建安，「Intranet 功能架構之研究」，第八屆國際資訊管理學術研討會論文集，頁 652-658。
- 【4】 陳仕昇，許秉瑜，陳彥良，「以可重覆序列挖掘網路瀏覽規則之研究」，資管評論，第九期，民國 88 年，頁 53-72。
- 【5】 陳建安，Intranet 功能架構與導入程序之研究，中山大學資訊管理研究所碩士論文，民國 86 年。
- 【6】 楊亨利、林俊宏，「組織構形與資訊科技構形的配合」，管理評論，第 16 卷第 2 期，民國 86 年 7 月，頁 49-96。
- 【7】 楊亨利、金士俊，「一個全球資訊網資料發掘的架構-以英語學為例」，資訊管理學報，第七卷，第二期，民 90 年 1 月，頁 143-181。
- 【8】 薛如芳，以歸納學習法自關聯式資料庫中發掘知識，交通大學資訊工程研究所碩士論文，民國 84 年。
- 【9】 Agrawal, R., Ghosh, S., Imielinski, T., Iyer, B. and Swami, A., "An Interval Classifier for Database Mining Applications," *Proceeding of the 18th VLDB Conference*, Vancouver, Canada, August, 1992, pp. 560-573.
- 【10】 Agrawal, R., Imielinski, T. and Swami, A., "Database Mining: A Performance Perspective," *IEEE Transactions on Knowledge* (5:6), December 1993, pp.914-925.
- 【11】 Agrawal, R. and Srikant, R., "Mining Sequential Patterns," *IEEE 11th International Conference on Data*

- Engineering*, Taipei, Taiwan, March, 1995.
- 【12】 Ahituv, N., Neumann, S., and Zviran, M., "Factors Affecting the Policy for Distributing Computing Resources," *MIS Quarterly*, December 1989, pp.389-401.
- 【13】 Bowman, M., Danzig, P. B., Manber, U. and Schwartz, M., "A Scalable, Customizable Discovery and Access System," *Technical Report CU-CS-732-94*, Department of CS, University of Colorado, Boulder, July, 1994.
- 【14】 Brass, D.J. and Burkhardt, M.E., "Centrality and Power in Organizations," in N. Nohria and R.G.(Eds.), *Networks and Organizations: Structure, Form, and Action*, Harvard Business School Press, Boston, Massachusetts, 1992, pp.191-215.
- 【15】 Chen, M. S., Han, J. and Yu, P. S. "Data Mining: An Overview from Database Perspective," *IEEE Transaction Knowledge and Data Engineering*, December, 1996, pp.886-883.
- 【16】 Chen, M. S., Park, J.S. and Yu, P. S. "Efficient Data Mining for Path Traversal Patterns," *IEEE Transaction on Knowledge and Data Engineering*, April 1998, pp. 209-221.
- 【17】 Cohen, W. M. and Levinthal, D. A., "Absorptive Capacity: A New Perspective On Learning and Innovation," *Administrative Science Quarterly* (35), 1990, pp.128-152.
- 【18】 Davenport, T.H. and Prusak, L., *Working Knowledge: How Organizations Manage What They Know*, Boston: Harvard Business School Press, 1998.
- 【19】 Desai, B. C., "Supporting Discovery in Virtual Libraries," *Journal of the American Society for Information Science* (48:3), Mar 1997, pp.190-204.
- 【20】 Dhar, V. and Tuzhilin, A., "Abstract-Driven Pattern Discovery in Databases," *IEEE Transaction Knowledge and Data Engineering* (5:6), December, 1993, pp.926-938.
- 【21】 Etzioni, O., "The World-Wide Web: Quagmire or Gold Mine?" *Communications of ACM*, (39:11), November, 1996, pp.65-68.
- 【22】 Fayyad, U.M., "Data Mining and Knowledge Discovery : Making Sense out of Data," *IEEE Expert* (11:5), October, 1996, pp.926-938.
- 【23】 Fiedler, Grover, and Teng, "An

- Empirically Derived Taxonomy of Information Technology Structure and Its Relationship to Organizational Structure, " *Journal of Management Information Systems*, Summer 1996 (13:1), pp. 9-34.
- 【24】 Frawley, W. J., Paitetsky-Shapiro, G. and Matheus C. J., "Knowledge Discovery in Databases : An Overview, " *Knowledge Discovery in Databases*, California, Edited by Paitetsky-Shapiro, G. and Frawley, W. J., AAAI/MIT Express, 1991, pp.1-30.
- 【25】 Friedman, R. and Krackhardt, D., "Social Capital and Career Mobility: A Structural Theory of Lower Returns to Education for Asian Employees," *Journal of Applied Behavioral Science*, (33: 3), 1997, pp.316-334.
- 【26】 Grupe, F. H. and Owrang, M. H. "Data Base Mining Discovering New Knowledge and Cooperative Advantage," *Information System Management*, (12: 4), Fall, 1995, pp.26-31.
- 【27】 Han, J., Cai, Y. and Cercone, N., "Knowledge Discovery in Databases : An Attribute-Oriented Approach," Proceeding of the 18th VLDB Conference, Canada, August, 1992, pp. 547-549.
- 【28】 Han, J., Cai, Y. and Cercone, N., "Data Driven Discovery of Quantitative Rules in Relation Databases," *IEEE Transaction Knowledge and Data Engineering* (5:1), February, 1993, pp.29-40.
- 【29】 Han, J., Y. Cai, N. Cercone, and Huang, Y." Discovery of Data Evolution Regularities in Large Databases," *Journal of Computer and Software Engineering*, 3(1), 1995, pp.41-69.
- 【30】 Han, J. and Fu, Y., "Discovery of Multiple-Level Association Rules from Large Databases," *Proc. of 1995 Int'l Conf. on Very Large Data Bases (VLDB'95)*, Zurich, Switzerland, September 1995, pp. 420-431.
- 【31】 Han, J., Koperski, K. and Adhikary, J., "Spatial Data Mining: and Challenges," *1996 SIGMOD'96 Workshop. on Research Issues on Data Mining and Knowledge Discovery (DMKD'96)*, Montreal, Canada, June 1996.
- 【32】 Han, J., Nishio, S. and Kawano, H., " Knowledge Discovery in Object-Oriented and Active Databases," *F. Fuchi and T. Yokoi (eds.), Knowledge Building and*

- Knowledge Sharing*, Ohmsha, Ltd. and IOS Press, 1994, pp. 221-230.
- 【33】 Han, J., Zaine, O. R., and Fu, Y., “Resource and Knowledge Discovery in Global Information Systems: A Scalable Multiple Layered Database Approach,” *Proc. of a Forum on Research and Technology Advances in Digital Libraries (ADL'95)*, McLean, Virginia, May 1995.
- 【34】 Helleloid, D. and Bernard, S., “Organizational Learning and a Firm’s Core Competence,” in G. Hemel G. & A. Heene (Eds.), *Competence-Based Competition*, New York: John Wiley, 1994.
- 【35】 Hendriks, P., “Why Share Knowledge? The Influence of ICT on Motivation for Knowledge Sharing,” *Knowledge and Process Management*, (6:2), 1999, pp.91-100,.
- 【36】 Houtsma, M. and Swami, A., “Set-Oriented Data Mining in Relational Databases,” *Data & Knowledge Engineering*, (17:3), September, 1995, pp.245-162.
- 【37】 Intranet group of NASA, “The Intranet White Paper-Definition, Benefits and Challenges,” 1995, Available from <http://pscinfo.pscni.nasa.gov/intranet/html/home.html/>.
- 【38】 Javenpaa, S.L. and Ives, B., “Organizing for Global Competition: The Fit of Information Technology,” *Decision Sciences* (24:3), 1992, pp.547-580.
- 【39】 Kalakota, R. and Whinston, A., *Electronic Commerce: A Manager’s Guide*, Addison-Wesley, 1997.
- 【40】 Leifer, R.P., “Matching Computer-Based Information Systems with Organizational Structures,” *MIS Quarterly* (12:1), March 1988, pp.63-73.
- 【41】 Mintzberg, H. “The Structuring of Organizations: A Synthesis of the Research,” Englewood Cliffs, NJ: Prentice-Hall,1979.
- 【42】 Nonaka, I. and Takeuchi, H., *The Knowledge-Creating Company*, Oxford University Press, 1995.
- 【43】 Pei, J., Han, J. Mortazavi-asl, B., and Zhu, H., “Mining Access Patterns Efficiently from Web Logs,” PAKDD’00 (Proc. Pacific-Asia Conference on Knowledge Discovery and Data Mining, Kyoto, Japan, April 2000).
- 【44】 Piatetsky-Shapiro, G., “Discovery, Analysis and Presentation of Strong Rules,” in G. Piatetsky-Spapiro and W.J.



- Frawley (Eds.), *Knowledge Discovery in Databases*, AAAI/MIT Press, 1991, pp. 229-248.
- 【45】 Raggett, D., “HTML 3.2 Reference Specification,” *W3C Organization*, 1997, Available from <http://www.w3c.org/TR/REC-html32.html/>.
- 【46】 Schwartz, M. F., Emtage, A., Kahle, B. and Neuman, B. C., “A Comparison of Internet Resource Discovery Approaches,” *Comput. Syst.* (5), Fall, 1992, pp. 461-493.
- 【47】 Siberschatz, A. and Tuzhilin, A., “On Subjective Measures of Interestingness in Knowledge Discovery,” *First International Conference on Knowledge Discovery and Data Mining*, August 1995.
- 【48】 Tellen, S. L., “The IntraNet Architecture: Managing information in the new Paradigm,” Available from <http://www.amdahl.com/doc/products/bsg/intra/infra.html/>.
- 【49】 Weibel, S., “A Proposed Convention for Embedding Metadata in HTML,” 1996, Available from <http://www.oclc.org:5046/~weibel/html-meta.html/>.
- 【50】 Weibel, S., Godby J., and Miller E., “OCLC/NCSA Metadata Workshop Report,” 1996, Available from [http://www.oclc.org:5046/oclc/research/...rences/metadata/dublin\\_core\\_report.html/](http://www.oclc.org:5046/oclc/research/...rences/metadata/dublin_core_report.html/).
- 【51】 Yoon, J. P. and Kerschberg, L., “A Framework for Knowledge Discovery and Evolution in Databases,” *IEEE Transactions on Knowledge and Data Engineering* (5:6), December, 1993, pp.973-979.
- 【52】 Zack, M.H., “Managing Codified Knowledge,” *Sloan Management Review* (40:4), Summer 1999, pp.45-58.
- 【53】 Zaine, O.R., *Resource and Knowledge Discovery from the Internet and Multimedia Repositories*, Ph.D., Simon Fraser University of Computer Science, 1999.
- 【54】 Zaine, O. R., Xin, M., and Han, J., “Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs,” Proc. Advances in Digital Libraries Conf. (ADL'98), Santa Barbara, CA, April 1998, pp. 19-29.